

Era of disappointment ends

Mrs Margaret Thatcher's remarkable eleven years as British Prime Minister, which end this week, may have enhanced Britain's reputation in many fields, but not in research. A spell of unconviction politics would be welcome.

WHAT will be the consequences for British science of Mrs Margaret Thatcher's precipitate departure from the British government this week? It is too soon to know. None of the three competitors for her job has much to say on the subject, although one, Mr Michael Heseltine, has spoken of the need to improve "training", measured by yardsticks of inter-European competitiveness, and of the need to win commercial benefits from still-high British spending on defence research and development. But the future cannot be worse than the past. Indeed, a new government will (or should) take the opportunity of mere change to put right some glaring errors — the calamity over university finances occasioned by the Universities Funding Council's abortive bidding system, for example.

An honest woman

It is curious that the first British prime minister with a technical qualification (a degree in chemistry from Oxford in 1947) should have become so closely associated with the decline of basic science in Britain. How can that have come about when, within a year of her election, Thatcher promised that the "science budget" (public spending on the publicly supported research councils) would be "protected" against inflation? Especially when she kept her word? Public spending under this heading has indeed increased from roughly £500 million in 1980 to £850 million now — an increase of some 10 per cent in real terms. What more could an honest woman have done?

Lots. The promise to protect the science budget was an illusion. Within a few months, it had been followed by an ill-considered, even thoughtless, reduction of university budgets on the grounds that "higher education must share in the general reduction of public expenditure". Committed as they were to students then on the rolls, all but the strongest universities promptly skimped on their support for research, making a nonsense of the supposedly hallowed "dual-support system".

The government, apparently indifferent to the consequences, then set about reducing the funds freely available for research in other ways. The nominally autonomous research councils were required to spend considerable sums on the government's pet causes, information technology for example. In the wake of the Falklands war, the budget for Antarctic research, previously shrinking, was substantially increased at the expense of

other spending. The Ministry of Agriculture, Fisheries and Food sent the then Agricultural Research Council (now the Agricultural and Food Research Council) into sharp and continuing decline by renegeing on the understanding, with a previous government in 1972, that the ministry would "commission" a substantial fraction of the research council's work. Paying people to retire early became the first call on what funds were left.

Although no one of these actions would, by itself, have been calamitous, their ostentatiously arbitrary concatenation, and the sense that there might be no limit to the indignities the research enterprise could be made to suffer, engendered a slump of researchers' morale. What else, when active researchers found their teaching loads increased and the rest of their time increasingly occupied by committees on institutional reorganization? Or when publicly supported laboratories found their ablest members prematurely put out to grass — or taking the government's early retirement shillings and finding creative jobs elsewhere?

Sadly, the managers of the British research enterprise must share some of the blame for the horrors of the past decade. Research council heads and the administrators of university funds have consistently been too compliant with the government's wishes, too ready to administer policies in whose genesis they played little part and under impossible constraints. Public protest has been conspicuous by its absence, on the effete British principle that the great and the good function most effectively in private dialogue with civil servants. The result is that the research community was given every reason to believe that its leaders had been captured by the government, which was often true. In the long decade now past, an occasional protest resignation would have done wonders.

Timorous management

The research council heads have also been timorous in the management of such of their own affairs as remained in their gift; they have been free with earnest declarations of the need to increase spending on unsolicited research grants, but have done less than they might to reduce spending on fixed commitments, in-house institutes and laboratories included. The largest of the research councils, the Science and Engineering Research Council, still spends only about a quarter of its funds on untied research grants. Who can wonder that it is now again in a jam?



Luckily, the new chairman, Sir Mark Richmond, is making the right noises (see page 377). He deserves the research community's full-throated support. The change of government should help him get his way.

These complaints may explain how the past decade has been made miserable for the British research enterprise, but not why that happened under a prime minister expected to be sympathetic. There have been two separate sources of stress, one of which stems from the legend that Thatcher is a "conviction politician" — the doctrine that belief is the determinant of policy, practical considerations notwithstanding. This, ironically, is the origin of the policy that seems chiefly to have made her seem an electoral liability to her own party; the scheme for raising local taxes by means of a fixed annual charge for every adult member of a community stems directly from Thatcher's conviction that such an arrangement *should* make local governments more directly accountable to their electorates, and thus less tempted to extravagance. The belief might have been proved correct if it had been possible peacefully to effect the change to the new system of taxation. All three of Thatcher's would-be successors agree that the new tax must be reformed.

Belief and prejudice

There are many examples of how belief, sometimes indistinguishable from prejudice, has marked the past decade's administration of research and higher education in Britain. The decision, at the end of 1980, that university budgets should be reduced over the succeeding three years, without contraction of capacity, stemmed from the conviction that academics are a bunch of layabouts. So, some years later, did the decision that academic tenure should be extinguished. (The formalities are still to be completed, but the practice has almost gone.) Stagnant research council budgets have been justified by the belief that restraint *should* impel them to decide on priorities. (Yet when it seemed, in 1986, that the research councils collectively might decide not to participate in the development of the new accelerator at CERN, the European high-energy physics laboratory, the government reserved for itself the final say.)

The belief that Britain would be more prosperous if researchers behaved as if they were industrialists, or at least made partnerships with industrial companies, has spawned a host of special schemes for research support, none of them conspicuously successful. The most recent elaboration of this belief, that government should not contribute to the cost of what is called "near-market" research (that seeming to offer opportunities for immediate exploitation), has created distracting dilemmas for both industrial and academic researchers and has become an imprecise justification for not spending money on research.

Successive Thatcher governments have nevertheless themselves been oddly timid when faced with the need for a research strategy. In 1982, the government rejected the argument of the House of Lords Select Committee on

Science and Technology that some member of the British cabinet should be given (among other functions) responsibility for research on the grounds that the prime minister, being a scientist, could handle that herself. The expectation at the time that she would have neither the time nor the detachment has been justified by events.

Much the same has happened to Thatcher's chairmanship of the grand new advisory council, called the Advisory Council on Science and Technology, which was meant to take a view on all issues touching research in Britain, but which has accomplished nothing worthwhile in the past two years. Its members, mostly industrialists, have been, like the prime minister, too busy to give it their attention. Nobody is surprised, merely disappointed. The best hope is that Thatcher's successor will create an independent advisory body functioning at least as well as the old Advisory Council on Scientific Policy until its abolition after the election of 1994.

For the British scientific enterprise, the past decade has been at best a disappointment. Excellent work continues even outside the handful of institutions emerging as centres of excellence, but there are fewer opportunities for British research groups to acquire the strength to be internationally competitive. For what was meant to be an enterprise culture, it is odd that young men and women with bright ideas should be the most commonly frustrated. Yet more senior people are too much distracted from research — institutional reorganization continues, as does a kind of computer-dating with industrial companies under the much-vaunted LINK scheme — and remain poorly paid as well. Two worries lie ahead. First, those who followed the government's exhortations of the mid-1980s to become entrepreneurs may be badly caught out by the worsened economic climate, with high inflation and high interest rates. Second, that that and the other events of the past decade will discourage the young.

A lot of money

The lacuna of understanding underlying the policies of the three successive Thatcher governments on research has been their failure to appreciate why they were spending money on any scale on basic research. In an inimitable declaration (in 1980), "£500 million is a lot of money". So, now, is £850 million. And the sum is even larger when augmented by the notional 30 per cent of university recurrent costs supposed still to support the academic half of the dual-support system, notionally £600 million a year at present. Sums like these are indeed "a lot" when they have no apparent purpose. Yet the Thatcher governments' were never able to regard research spending as an essential adjunct of the process of training young people in research. Whence the incessant empty clamour for industrial applicability. Nobody in Britain will now ask that Thatcher's departure signals a return to an ivory tower, or that the purse-strings will dramatically be loosened. But a more rounded understanding of what basic research is for, and what it is like, would be a great boon. □

SERC cuts hard to avert £40 million cash crisis

- Research grants, studentships frozen
- 'Big science' projects under review

London & Cambridge

British science was last week plunged into crisis as the Science and Engineering Research Council (SERC), the largest funding body for UK civil science, postponed its current round of research grants and studentships, and placed a freeze on staff appointments, to avoid a predicted shortfall of £40 million in 1991–92. Large projects agreed in principle, but not yet funded, including an 8-metre optical telescope and a gravity wave observatory, now seem certain to be delayed.

SERC's financial crisis, like that which has emerged at the Medical Research Council (MRC) in recent weeks (see *Nature* 348, 184; 15 November 1990), is blamed largely on higher-than-expected inflation and pay settlements. The disclosure of the council's dire financial position follows the confirmation that the Agricultural and Food Research Council (AFRC) will shed some 380 research staff by the end of the financial year as part of its continuing restructuring, and the University of Bristol's decision to freeze new appointments and equipment purchase, to reduce an estimated deficit of £4 million in the current year.

A small group of council members will meet over the coming months to consider further measures needed to resolve SERC's finances. But the council's new chairman, Sir Mark Richmond, visiting Cambridge last Friday (23 November) with Education and Science Secretary Mr Kenneth Clarke to open a 32-metre radiotelescope, also promised a longer-term broad review of SERC's activities. Richmond wants to reduce spending tied up in 'big science' projects and international collaborations so that small research grants and studentships can be expanded.

Richmond used the telescope opening to attack government policy on science and the universities. The current university funding crisis (see *Nature* 348, 3; 1 November 1990) could "ring the death knell" for the universities' support for large science projects, he said. (The University of Manchester, where Richmond was vice-chancellor until October, currently provides £800,000 a year for radioastronomy at Jodrell Bank, he pointed out.) SERC's plans to fund an 8-metre optical telescope have been sent "sprawling" by the disappointingly low science budget for 1991–92 (see *Nature* 348, 183; 15 November 1990), Richmond added.

Clarke sought to absolve himself from blame by pointing out that he had recently "inherited" the government's science spending plans, after his transfer from the Department of Health, but repeated the claim that the 1991–92 budget is a "real-terms level settlement". After the ceremony, Richmond said that inflation is certain to mean a cut in the amount of science that can be supported: "The Secretary of State is wrong".

Richmond declined to speculate in a "knee-jerk reaction" whether his long-term review will cut back SERC's own laboratories to release money for research grants. But he said that SERC must negotiate tougher agreements to control spending on large international projects. The science budgets of CERN (the European particle physics centre) and the European Space Agency will be discussed in December. Britain's contributions to both are paid through SERC's budget, and fluctuations in international exchange rates can squeeze SERC's domestic spending. In 1987, the problem was so severe that Britain came very close to pulling out of CERN.

Frequent reviews of grants to SERC's Interdisciplinary Research Centres (IRCs), introduced under the previous chairman, Sir William Mitchell, to provide a focus for work that may eventually yield commercial results, are also likely. Three new IRC grants, totalling some £18 million, are among the projects now in abeyance. Richmond said that he is unhappy with the present arrangement, where a new IRC can receive guaranteed funding for six years. He also hinted that SERC may introduce a rigid cash limit on all research grants, rather than linking their value to inflation.

But the most important change in the expected upheaval within SERC is in the balance between spending on 'big science' projects, such as astronomy and high-energy physics, and smaller research grants. Unlike his predecessors, Richmond comes from a 'small science' background in biology. "Big old projects" will have to be cut if "big new projects" are to be started, Richmond said.

During the summer, the Advisory Board for the Research Councils suggested to Mr John MacGregor, the previous Education and Science Secretary, that SERC's support for 'big science' should be reviewed. But this is thought to have been opposed by the then Mitchell-led SERC.

The other research councils have no plans to follow SERC and the MRC and suspend their current grant rounds. Professor John Knill, chairman of the Natural Environment Research Council (NERC), says that NERC's financial management is such that it would never find itself in that position.

But SERC staff say that the council's current financial plight should not be taken as evidence of lax management. To a greater degree than for the other research councils, the vast majority of SERC-supported scientists are in the universities, rather than research council laboratories. This means that SERC finds it very difficult to account for future pay settlements in its plans.

Peter Aldous

OZONE DESTRUCTION

Methyl chloride gas implicated

Washington

To the list of human activities that destroy the ozone layer can now be added lighting a campfire. 'Biomass burning' — the combustion of wood, straw or other natural fuel — releases large amounts of methyl chloride, which, as a chlorine-based gas, has many of the ozone-attacking properties of the well-documented chlorofluorocarbons. Human activity, mostly in the developing countries, pumps some 2–5 million tonnes of methyl chloride into the atmosphere each year, accounting for 5 per cent of the total concentration of ozone-



destroying chemicals in the troposphere, according to a recent report from the Maryland-based Institute for Energy and Environmental Research.

Although the chlorofluorocarbons-eliminating Montreal Protocol will force the largest reduction in ozone-destroying gases, its priorities are chiefly focused on emissions that come from the developed countries, says Arjun Makhijani, one of the authors of the report. In the developing world, methyl chloride represents some 26 per cent of total ozone-unfriendly emissions.

Christopher Anderson

First European experiment

Munich

A TEAM of Italian researchers is expected to receive permission for the first human gene therapy experiment in Europe before the end of the year, setting the stage for further experiments in other countries.

Claudio Bordignon of the Istituto Scientifico San Raffaele in Milan is the leader of the team and was an associate of the group that last July won approval for a similar experiment from the Recombinant DNA Advisory Committee (RAC) of the US National Institutes of Health (NIH; see *Nature* 346, 402; 1990). Bordignon says that there are groups in the Netherlands and France that will soon be ready to perform gene therapy experiments.

There has been an outcry against the experiment in Italy, because of fears that unscrupulous researchers may visit Italy to perform experiments on humans. A group of parliamentarians asked the Health Ministry for a moratorium on gene therapies, especially germ-line therapy. But Francesco de Lorenzo, the Health Minister, says that it is up to the hospital in question to decide if the experiment meets medical and ethical standards.

Bordignon expects speedy approval for the experiment because his protocol is more conservative than that approved by RAC in the United States. He is treating a three-year-old child for a rare inherited immune disorder caused by a deficiency of the enzyme adenosine deaminase (ADA). The deficiency causes an almost complete absence of normal immune function, meaning that children who suffer from it — there are fewer than 20 worldwide — must live in a sterile 'bubble' unless the enzyme can be replaced.



Bordignon intends to remove some of the child's T lymphocytes and insert a normal human ADA gene into them. But the altered T-cells will be put back into the child's body only if current treatment with bovine enzyme fails. Until that time, the altered T cells will be frozen.

Unlike the United States, European countries have so far viewed somatic gene therapy as just another medical treatment that does not present special moral or ethi-

cal problems. There is no equivalent to RAC in Italy or many other European countries. Nevertheless, opposition to genetic engineering may make such experiments difficult in countries such as Germany. Peter Lange, head of the section for basic questions in biology and health at the German Research and Technology Ministry says that "as soon as the press finds out about this [in Germany], there will be a huge discussion. It has been so quiet so far", he adds, "only because the public does not realize how close [the experiments] are to reality."

The EC consensus, reiterated at a March 1990 meeting of research ministers at Kronberg in Germany, breaks down when it comes to germ-line gene therapies. Although the ministers agreed to reject germ-line therapies for the moment, says Lange, their opinions diverged about whether such therapies will be permitted in their countries in the future. Changing the germ line using genetic engineering will become illegal in Ger-

COMPUTER NETWORKS

Eastern Europe comes online

Budapest

ACADEMIC computer users in Eastern Europe will be the among the first to benefit as foreign manufacturers rush to fill the vacuum in the market for computers. Following a pattern set in the early 1980s in North America and Western Europe, IBM and other companies are offering their equipment to universities at little or no cost in the hope of persuading a generation of students and professors of the value of their products.

The scientific community sees the computer technology as essential to raising its standards. "We hope we can catch up ten years in just two or three years using these new computers", says Alexander Szalay, an astrophysicist and network manager at Loránd Eötvös University in Budapest.

The most ambitious project so far is the Academic Computer Initiative of IBM Europe under which it is installing mainframe computers (the largest computers ever installed in the region) in Czechoslovakia, Hungary, Yugoslavia and Poland. The countries will then be linked to the West through the Vienna node of the European Academic Research Network.

On 20 November, IBM signed a cooperative agreement with three universities in Budapest under which the company will provide a mainframe computer and Hungary's first-ever fibre optic network.

The Budapest and Belgrade computers will be installed next spring, says Judit

many at the beginning of next year.

The EC Commission is setting up a committee to look into the ethical and social implications of human genome analysis (which may also deal with gene therapies) but an EC official in Brussels says that it is up to the member states to regulate the therapies at a national level. In the official's opinion, any attempts to ban somatic gene therapy throughout the EC might well fail because of Article 36 of the Treaty of Rome, which allows nations to deviate from Community policy with respect to issues of "public morality . . . public safety and the protection of human life and health".

The EC official also fears the effect of press coverage on this field of research in Europe. The official says that "merely drawing attention to this situation creates what looks like a legislative vacuum . . . that politicians will rush to fill", with the consequence that an inflexible bureaucracy would be created "with an interest in self-perpetuation". It would be far better, says the official, to let researchers themselves advise the EC how to proceed in such a fast-moving field.

Steven Dickman

Quittner of IBM Hungary. Earlier this month another computer of the same model, a 3090-120, was installed in Prague. At the same time, Digital Equipment Corporation (DEC) is about to sign an agreement with Eötvös University to install, at large discounts, both hardware and software for electronic mail within the university and, eventually, among several universities in Hungary.

According to George Balatoni of DEC Hungary, the size of the potential Eastern European market should not be underestimated. Balatoni estimates DEC will sell \$100 to \$150 million of equipment in Eastern Europe in the next five years, beginning with Hungary, the only country so far where DEC has established a majority-owned subsidiary.

Obtaining both IBM and DEC products at the same time is an especially useful combination for researchers in Hungary, says Szalay. Until now, owing in part to Western high-technology export restrictions, the most advanced computers in the country were microcomputers. IBM will fill a gap with its large computers, while the DEC systems will be able to connect systems of many different manufacturers, including IBM and Apple.

Szalay praised both companies for "listening to our needs" and helping Hungary to choose not the latest technology but rather the most cost-effective system that it will be able to afford to expand.

Steven Dickman

PARTICLE PHYSICS

CERN staff in pay dispute

London

STAFF at CERN, the European particle physics centre in Geneva, are calling for the resignation of their director-general, the Nobel laureate Professor Carlo Rubbia, in a deepening dispute over pay and pensions. Rubbia has been singled out because a pay package he proposed together with the staff association over the summer has been rejected by an advisory committee to CERN's governing council.

CERN staff and Rubbia proposed a 6.4 per cent pay increase in real terms by 1993, but Rubbia says that CERN member states may be unable to commit themselves to pay increases beyond 1991. The CERN council is expected to offer just a 3 per cent increase in the coming year, when it meets on 14 December — with no longer-term promises. Dr Henry Atherton, an applied physicist and vice-president of the staff association, says that CERN has also allowed a debt of some SFr350 million (about £140 million) to build up in the staff pension fund, by failing to meet contributions which were increased in the early 1980s. Money has been withheld from the fund to pay for CERN's Large Electron-Positron collider, Atherton says.

The call for Rubbia's departure is little more than a posture on the part of the staff association, but CERN staff have also started a work-to-rule. Support staff and a small number of research physicists are refusing to use their own cars on the sprawling CERN site or to work unpaid overtime, and have abandoned the use of CERN's internal electronic communications in favour of handwritten notes. Atherton says that CERN staff will now obey strict safety rules to the letter.

So far, the dispute has not strongly affected scientific output, according to Dr David Miller, a regular user of CERN's facilities from University College London. Others suggest that the squabble is diverting attention from the real problem: the reluctance of members to increase contributions to fund new science.

There are also doubts among CERN's outside users over the validity of the staff pay claim. CERN staff argue that their pay lags behind that in comparable international organizations (such as the European Southern Observatory) and in local Geneva industry. But Professor Don Perkins from the University of Oxford a former chairman of the UK Science and Engineering Research Council's Nuclear Physics Board, says that a better comparison is with outside users, who are mostly paid much less than CERN staff.

Rubbia says that the pensions issue is not an immediate problem, because CERN is legally obliged to meet its contri-

butions. Payments have lagged behind the pension fund's total agreed value by 19 per cent, Rubbia says, but this is within the limits allowed by Swiss law.

But CERN member states are unlikely to guarantee pay increases as far forward as 1992-93, because national science budgets are fixed on a yearly basis. Rubbia says that in the absence of an agreed total CERN budget for 1992-93, individual components such as pay increases cannot be guaranteed, but adds that he will resist "very strongly" any attempts to reduce spending on the running costs of CERN experiments. This will simply shift the problem onto member states' domestic science budgets, he says, if CERN's outside users are to continue with the present volume of work. **Peter Aldhous**

ASTRONOMY

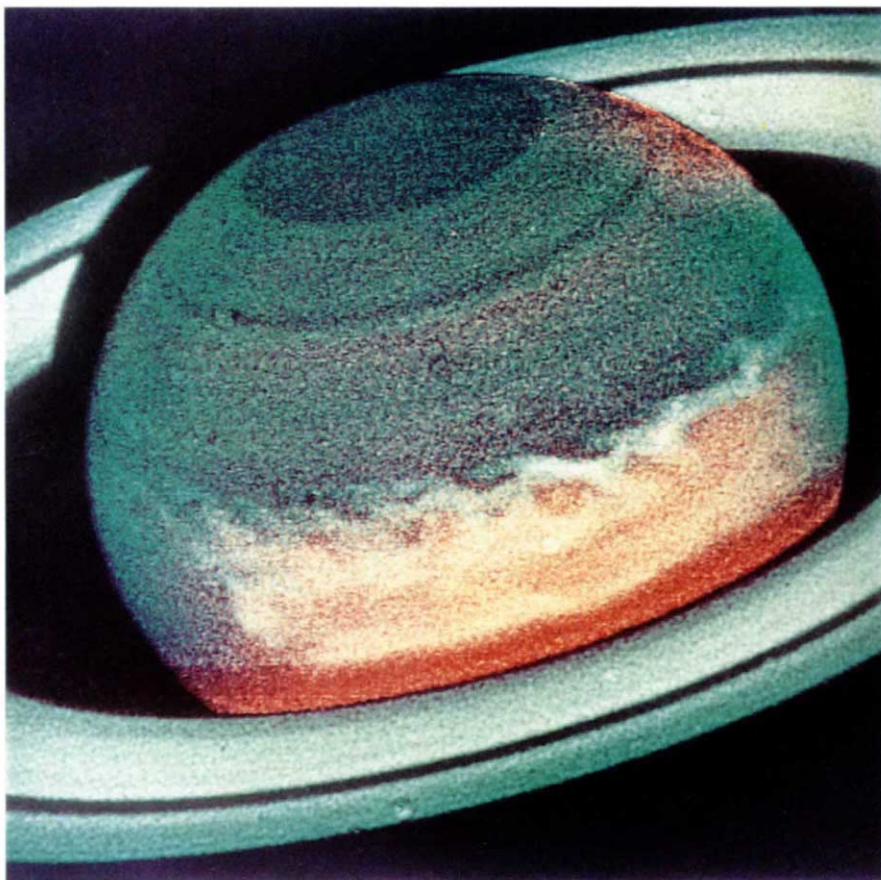
European society seeks a role

London

MORE than 600 astronomers have come together to form the European Astronomical Society (EAS), the first European-wide professional body in the field. The society will fulfil a useful role if it can provide a forum for the discussion of collaborative European astronomy projects, British astronomers say. Some feel that important collaborations such as the European Southern Observatory could be better coordinated.

Britain already has an analogue to the new society in the Royal Astronomical Society (RAS). But Dr Paul Murdin, from the Royal Greenwich Observatory, says that the new EAS should still attract UK members. **Peter Aldhous**

Hubble spotlight on Saturn's storms



Saturn's Great White Spot, a massive storm in the planet's hydrogen- and helium-rich atmosphere, has astounded astronomers since it first developed two months ago. Almost certainly a result of upwelling traced by ammonia-ice clouds from the lower atmosphere, the spot has grown from 20,000 km across in early October to encompass much of Saturn's equator now. To learn the most about this nearly unique event (only four such spots have been seen before, in 1876, 1903, 1933 and 1960, all smaller than the present one) NASA turned the planetary camera of the Hubble Space Telescope towards it on 9 November. The high resolution available from the telescope reveals unprecedented detail, showing in particular the evidence of turbulent eddies around Saturn's spot that resemble those around its larger cousin, Jupiter's Great Red Spot, which has lasted at least 400 years continuously. (Blue in the picture indicates low-level clouds; red, high-level clouds.) □

Ape import cuts hit research

San Francisco

THE eight-month-old restriction on the import of some of the most important research primates has tripled the price of the animals and seriously curtailed primate research in the United States, according to researchers.

The shortage of research primates is growing as primate importers across the United States come into line with the strict quarantine requirements demanded by the federal Centers for Disease Control (CDC) earlier this year after the discovery of a suspicious filovirus that was making ill and killing imported monkeys and apparently infecting handlers as well.

Researchers say they must now endure long delays to get the primates they need for research. "There's no question it's had a definite negative impact on research and the whole supply line of research animals into the United States", says Jeff Roberts, assistant director of the California Primate Research Center at the University of California at Davis.

Many research projects that rely on imported monkeys and apes have been halved, he estimates, because the required animals are either not available or are now prohibitively expensive. The cost of one cynomolgus macaque, for example, has risen from \$400 to nearly \$1,200 in recent months.

The primatologists' pinch is the result of a year of nervous manoeuvring and rule-writing. Last autumn, in several shipments of monkeys from Asia to the United States, a filovirus was identified that appeared quite similar to Ebola, a virus that caused large outbreaks of often-fatal human disease in Africa in the 1970s. When testing of humans who had had contact with the imported animals revealed several positive antibody tests, the CDC took action.

In March they issued a mandatory series of guidelines on quarantine and handling of primates. The CDC also told US primate centres that they would begin unannounced site inspections. As a result, many primate importers, including three of the country's largest centres, which together account for more than 80 per cent of all US imports, were closed down last spring.

Meanwhile, the New York State Health Department added an additional layer of safeguards by declaring that it would permit cynomolgus, rhesus or African green monkeys to enter the state only if they had been quarantined in their country of origin for 60 days and had tested negative for the filovirus. Even so, animals entering the United States through New York, as do most of those imported into the country, must be quarantined and re-tested after importation. Importers must

also make special arrangements with wary airlines to ensure that employees are not exposed to the virus during transport.

Despite the new restrictions, the flow of primates into US institutions is at last beginning to pick up again. In April, the CDC announced that special permits would be made available for the importation of cynomolgus. African green and rhesus monkeys if importers met certain requirements. Since then, primate importers, university laboratories and zoos around the nation have been scrambling to remodel quarantine facilities and to develop stringent new procedures for disinfecting facilities, handling animals and waste and routinely testing animals and human handlers for the filovirus. In recent weeks, several importers have been granted special permits and a handful have been given rights to unlimited importation for a 180-day period. Many others are in the process of applying to the CDC for reinstatement.

In California, where 20 of the state's former 36 primate importers are under review for new permits, still another layer of bureaucracy and another round of quarantine and filovirus testing is needed for imported primates. For 20 years, the state has required a 30-day quarantine for incoming primates, even those coming from other states. That policy remains in place, with recent upgrades, despite its redundancy with the federal requirements.

Despite the problems caused by shortages and the fact that none of the human infections have so far resulted in any symptoms, the "need for extreme caution" remains, according to the CDC's Brian Mahy, who directs the division studying the filovirus infection.

The key question now, he believes, is whether the virus survives after infection in animals that are not killed by the pathogen. The existing evidence is that the virus does not survive long-term in the body, he says.

Those who use primates agree that it is crucial to protect their employees and their animals from the virus, but the CDC's response raises questions for many. Importers point out that the United States is the only importing nation to tighten regulations because of the filovirus, and no health problems have been reported in other countries. Meanwhile, the delays mean a sharp decline in the US breeding stocks of primates that help to sustain research institutions such as the seven regional primate centres sponsored by the National Institutes of Health. Zoos are also worried, because additional delays may begin to hurt their cooperative programmes to breed endangered species.

Elizabeth Schaefer

Maternal transmission at the zoo?

London

AN antelope from London Zoo, never fed the animal feed thought to harbour the infective agent of bovine spongiform encephalopathy (BSE), is being examined to determine if it died of the disease. If the 18-month-old kudu, which suffered from an unspecified neurological complaint, is found to have had a spongiform encephalopathy, fears that BSE can be passed down from mother to offspring will be reawakened: the animal's mother died of a spongiform encephalopathy in 1989 (see *Nature* 344, 183; 1990).

Maternal transmission occurs in sheep infected with scrapie, which is thought to be caused by the same as-yet-unidentified agent. More than 300 calves of BSE-infected cows are being studied by agriculture ministry epidemiologists, but none of the herd, now more than two years old, have yet developed the disease. Nevertheless, maternal transmission in an antelope would increase pressure for a breeding ban from the offspring of BSE cases.

The new scare comes as the government announced an inquiry into the animal feed industry, and banned the splitting of cow heads in abattoirs, to prevent aerosols of possibly infective brain tissue being sprayed over meat for human consumption.

A diagnosis for the new antelope case is expected within the next few weeks, after pathologists at the Central Veterinary Laboratory have examined the animal's brain. But James Kirkwood, senior veterinary officer at London Zoo, says that he "wouldn't be surprised if it turns out to be negative". Several suspected cases of maternally transmitted BSE in cattle have proved to be false alarms. Peter Aldhous

AIDS

Soviet television to raise AIDS funds

London

SOVIET television is planning an all-day broadcast on World AIDS day, 1 December, to raise money for AIDS research, modelled on the charity fundraising 'telethons' that have become a feature of British television in recent years.

This year's World AIDS day, co-ordinated by the World Health Organization (WHO) to highlight efforts to fight the worldwide pandemic, takes the theme of 'Women and AIDS'. WHO estimates that about 200,000 women will develop the disease within the next year. Heterosexual transmission of the disease to women is an increasing problem — 27 per cent of European women with AIDS are thought to have been infected through sexual intercourse with a HIV-positive man.

Peter Aldhous

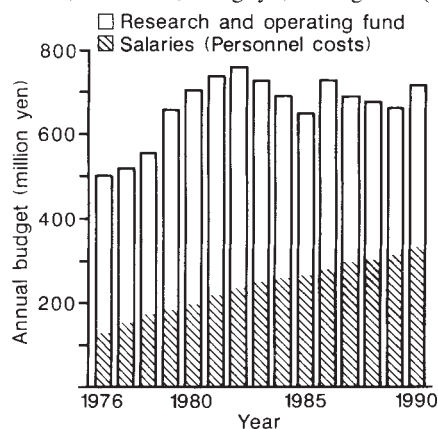
MITI yens for regional aid

Nagoya, Hiroshima & Kyushu

JAPAN'S Ministry of International Trade and Industry (MITI) is about to throw a few peanuts to its regional research institutes far from Tokyo in the hope of stimulating joint research between local industry and government.

Despite the boom in research and development in the private sector, Japan's government laboratories have suffered a financial squeeze for the past decade.

Built in the period of rapid industrial growth between the 1950s and the 1970s, MITI's regional laboratories in Hokkaido, Tohoku, Nagoya, Chugoku (a



region around Hiroshima) and Kyushu at first enjoyed quite rapid growth, both in money and people. But in the early 1980s, as the national budget suffered due to huge spending on public works — on roads and railways for bullet trains, for example — the Ministry of Finance clamped down on government spending. Government appointments (those of researchers included) were frozen, as were the general accounts for research and development of science-related ministries and agencies. In MITI's case, general account spending was forced down by a fixed percentage every year.

The result? Stagnation. The budget for MITI's Government Industrial Research Institute in Chugoku (GIRIC) shows what has happened (see figure). After growing at an average annual rate of more than 10 per cent until the early 1980s, the institute's total budget has since remained a level ¥700 million (\$5.4 million). Staff numbers have been almost constant at just over 50. But, with salary inflation, salaries have been eating up an increasingly larger share of the budget — nearly 50 per cent in 1990 compared with just over 25 per cent in 1976.

MITI has tried to cope with the steady erosion of its general account by drawing on its special account budgets, which are derived from taxes on oil and electricity and which are not subject to Ministry of Finance ceilings. While MITI's general account for research and development has

fallen from ¥85,000 million (\$670 million at current exchange rate) in 1980 to ¥68,492 million (\$540 million) this year, the special accounts have more than doubled, from ¥72,435 million to ¥153,140 million (\$1,206 million), in the same period.

MITI has also won extra funds of ¥26,000 million a year by the establishment of the Japan Key Technology Centre, funded by dividends of government-held shares of the recently privatized domestic telecommunications company, Nippon Telegraph and Telephone. But little of this money has filtered through to MITI's regional laboratories, still mostly dependent on the general account.

MITI has also had some success, in these hard times, in herding Japanese companies together into research consortia that work in collaboration with MITI institutes but are fed with only minimal amounts of government money. These include MITI's optoelectronic research projects and the recently established superconductivity research consortium headed by Shoji Tanaka of Tokyo University. But only MITI's larger laboratories near Tokyo, such as the Electrotechnical Laboratory (ETL) in Tsukuba, have gained from these ventures.

This may be about to change. MITI's budget request for next fiscal year includes a tiny request for ¥15 million (\$115,000) under the category of "regional largescale projects". Hitherto, such funds have been used only by MITI's regional research institutes, but the new money is earmarked for private industry in the vicinity of MITI's regional institute at Nagoya, and is intended to stimulate the already well-developed government/industry research on advanced ceramics in that area (see *Nature* 348, 186; 1990).

Katsuhiko Umehara, senior deputy director of MITI's Agency of Industrial Science and Technology (AIST), says that MITI now aims to have at least one such project for each of its seven regional laboratories. The laboratories will act as regional coordinators of the research.

Although the amounts of money involved in this initiative are minuscule, about \$1 million a year at most, one of MITI's traditional skills is that of goading industry into action with limited government funds.

A prime mover behind the new initiative is Katsuya Nakayama, director-general of GIRIC near Hiroshima. He has recently organized several symposia to introduce academics and industrialists in the Chugoku area to new MITI initiatives such as the Human Frontier Science Programme and the Intelligent Manufacturing System (IMS) project.

David Swinbanks

Controlling the tides

Hiroshima

DESPITE its shoestring budget, MITI's Government Industrial Research Institute, Chugoku (GIRIC), near Hiroshima, has succeeded in building one of the largest hydraulic models in the world with its 230-metre-long scale model of the Seto Inland Sea. Built at a cost of ¥1,600 million (\$12 million), the model comes complete with artificial tides, currents and 73 rivers which together replicate precisely conditions in



Monitoring the path of pollution.

the real sea which lies between the mainland Japan and the island of Shikoku.

One of the aims of the model is to come up with ways of artificially modifying water flow in order to reduce the severe pollution caused by industry, agriculture and waste from the large cities of Osaka and Hiroshima which border the sea. To provide effective answers from a scale model, fluctuations of the water surface have to be controlled to an accuracy of about a millimetre. A laser system is used that continuously feeds water-level readings back to a computer that commands a huge hydraulically operated system of weirs. A complete annual cycle of tides can be modelled in just 55 hours.

After completion in 1973, it took several years work by 25 researchers to fine-tune the model. Now that it is performing well, GIRIC researchers are experimenting by removing islands or creating new channels to find out the effects on flow. By introducing coloured water, they can monitor the path of pollutants. In one experiment, researchers at GIRIC have created a channel leading out of Osaka Bay, a site of stagnant flow and major pollution, to see if pollutants can be flushed out into the Pacific Ocean.

David Swinbanks

Drug debate expands

Washington & Paris

A US congressman last week accused the French drug manufacturer Roussel-Uclaf and US government officials of giving in to pressure from anti-abortion activists and withholding the abortion drug RU-486 from researchers studying cancer and other diseases.

Although initial hopes for RU-486 as the perfect at-home abortifacient have dimmed as the realities of its use became apparent (several clinical visits and companion treatments are required to use it safely and effectively), it has become a promising candidate for other medical applications. At a hearing last week of the subcommittee on regulation of the US House of Representatives Small Business Committee, researchers testified that RU-486 therapy could result in treatments for breast cancer, a rare glandular disorder known as Cushing's syndrome, other hormonal disorders and perhaps even hypertension and AIDS.

But because the Food and Drug Administration (FDA) imposed a ban last year on the commercial import of the drug, and Roussel-Uclaf has been reluctant to provide it to researchers, much of the research on non-abortion uses for RU-486 has ground to a halt, said subcommittee chairman Ron Wyden (Democrat, Oregon). "As a result of FDA's RU-486 policies, anti-abortion politics, and the manufacturer's business decisions, Amer-

UK RESEARCH COUNCILS

AFRC looks overseas

London

BASKING in the glow of a 44 per cent increase in funding from the European Communities this year, the UK Agricultural and Food Research Council (AFRC) has gone multilingual in its annual report, for the first time including summaries in French, German and Spanish. The change is not just cosmetic, says acting AFRC secretary Brian Jamieson. The council is expanding international collaboration at all levels, including increased involvement of North American and European scientists in its peer-review process, he says.

The other research councils have no plans to follow the AFRC's multilingual approach in their annual reports, but stress that they are nevertheless committed to increasing earnings from overseas. Annual reports are aimed primarily at the Secretary of State for Education and Science, a Natural Environment Research Council (NERC) spokesman says, and NERC uses carefully targeted brochures to market its science to overseas customers. Peter Aldhous

icans now suffering from horrible illnesses will not have the hope of cures they had before the import ban was imposed", Wyden charged.

Several researchers testified that they had been denied the drug by FDA or Roussel-Uclaf, or told by the company that sufficient supplies of RU-486 could not be guaranteed. In a memorandum to subcommittee members, Wyden said that Roussel-Uclaf officials had said that they were not willing to risk opposing the US government on distribution of the drug for fear that an anti-abortion backlash could endanger worldwide RU-486 approval by the World Health Organization, which depends heavily on US financial support.

But FDA officials said that the agency had intended only to thwart personal use — not legitimate clinical research — and that including researchers in the ban had been a mistake. Once a scientist has obtained special FDA approval — as with any drug that has not been approved for commercial distribution — "getting the drug becomes the responsibility of the researcher and the drug company", says FDA spokesman Jeff Nesbit. And although no Roussel-Uclaf representative testified at the hearing, a company official said in an interview that it, too, has no intention of obstructing good research, as long as it is not related to abortion.

"So long as the protocol appears interesting to our scientific committee and fits in with our own developments, we will supply the drug", said Ariel Mouttet, marketing director for RU-486 at Roussel-Uclaf in Paris.

In the case of one US scientist, Katherine Horwitz, a breast cancer researcher at the University of Colorado who testified that she had been denied permission by the FDA to import the drug for a clinical trial, Mouttet said, "If [she] writes to us with a protocol and asks for RU-486, and our scientific committee decides it is an interesting study, she will get the drug". When researchers have been denied the drug, she said, it was because their protocols had been found scientifically unworthy, or were linked to abortion research.

George Chrousos, a researcher at the National Institutes of Health who is studying Cushing's syndrome, says he is now more optimistic that he will be able to secure enough RU-486 from Roussel-Uclaf to allow his research to continue. Although he had previously received enough of the drug to continue his work for several years, he expects to run out in a few months without a new supply. Recently, he says, he has been unable to get an assurance from the company that it would make RU-486 available to him. "It's been quite uncertain. They have not

refused, but they haven't given an answer", he says. But public outcry over the endangered research may have shifted the balance, he says.

Mouttet said that Roussel-Uclaf has no plans to commercialize the drug in the United States, or to allow clinical trials of the drug as an abortifacient to go ahead. (It still cannot be bought, even in France, for personal use of any kind.) But the company is taking part in US clinical trials where RU-486 is used in the treatment of endometriosis and meningioma, she said.

Mouttet also expressed surprise at the claims that are being made for RU-486 in breast cancer research. "We are only at the clinical research stage. There is absolutely no evidence that the drug is effective in such cases." She said it was "scandalous" to suggest that RU-486 could be used in the treatment of AIDS. "There is absolutely no evidence, not even theoretical". Such claims only served to raise the hopes of the sick "and to make Roussel-Uclaf look wicked", she said.

Meanwhile, Roussel-Uclaf is maintaining its "step by step" approach to commercializing the drug as an abortifacient elsewhere, Mouttet said. A decision by the UK Department of Health is expected in January, following an application made this year. If Britain agrees to authorize the drug, Roussel-Uclaf will probably target Scandinavia, where abortion is legal and noncontroversial. So far, 50,000 pregnancies have been terminated by using RU-486.

Christopher Anderson & Peter Coles

SCIENTIFIC PUBLISHING

Coping with copyright

London

To ease the minds of its US customers, who must obey strict copyright laws, the Yorkshire-based British Library Document Supply Centre (BLDSC) has decided to charge a copyright fee on top of its normal duplication costs for scientific manuscripts. The library, which is the world's largest supplier of photocopies of articles from scientific journals, says the move will help to clear up what has been "an uneasy relationship" with scientific publishers on the copyright issue.

The new scheme, which will start in April 1991, is not compulsory — scientists will be exempt from the charge if they sign, as at present, a declaration that the copy is for personal use only. But those paying the copyright fee (about £1 per paper) will be able to distribute copies more widely. US drug companies, among BLDSC's most important customers, support the move, as it will allow them to store copies of important research papers in their own libraries, or pass copies around large research groups, says Graham Cornish, the library's copyright officer. Peter Aldhous

RISK ASSESSMENT

Alar study triggers lawsuit

Washington

A group of US apple-growers plans to file a \$250-million lawsuit tomorrow against the Natural Resources Defence Council (NRDC) and television network CBS for publicizing a study asserting that Alar, a pesticide used on apples, was a powerful carcinogen.

Charging "product disparagement", the farmers claim that the 1989 NRDC report, *Intolerable Risk: Pesticides in our Children's Food*, was based on flawed science and was not properly peer-reviewed. The study cited in the report had been carried out for the US Environmental Protection Agency, but had been rejected by the agency, says J. Jarrette Sandlin, a lawyer from Yakima, Washington, who represents the farmers. The lawsuit claims that, among other errors, the study exceeded reasonable levels of Alar concentration in determining the toxicity of the growth-regulating chemical. The study "purposely violated standards of laboratory science", to link Alar with cancer, says Sandlin. "The whole case will hinge on whether the report was a legitimate scientific analysis of Alar."

NRDC spokesman Paul Allen says that the organization stands by the report, but a detailed response will have to wait until the NRDC lawyers have seen the lawsuit. He says, however, that the report was peer-reviewed by nearly a dozen independent scientists.

Lawyers for both sides expect that the

case will test the scientific interpretation of the first amendment to the US constitution, the right to free speech. Scientists hold dear their right to be wrong without fear of legal retribution. Should this case go to a jury trial, 12 non-scientists could be determining the scientific validity of the study and NRDC's conclusions, a prospect that many researchers understandably find frightening.

Thomas Krattenmaker, a first-amendment expert at Georgetown Law School, says that although scientific studies are afforded no special protection under the law, the apple-growers would probably have to prove that NRDC showed a malicious disregard for the truth, leading to a demonstrable harm, to win the case.

Although the growers say that apples to a value of between \$200 and \$250 million were lost in the wake of the report and the CBS coverage, Krattenmaker suggests that they will have to prove that it was the study alone, rather than the facts in the study, that caused consumers to abandon red apples.

"If [the conclusion of the report] was just a mistake, we understand", says Sandlin. "But the apple-growers are asking the jury to make a decision on what is the outer boundary of such a statement. Did they act in reckless disregard of the truth? The law does not let a group lie to 40 million people [the CBS audience] and get away with it."

Christopher Anderson

BIOTECHNOLOGY INDUSTRY

Mum's the word . . .

Washington

WORD has it that the latest fashion at the biotechnology company Genentech, Inc. is T-shirts that read "Shhh, I have something to tell you . . . But don't tell Mollie."

"Mollie" is Mollie Raab, wife of G. Kirk Raab, president and chief executive officer of Genentech, who was fined \$162,400 by the Securities and Exchange Commission (SEC) last week for giving her brother price-sensitive information about the merger of Genentech and health-care conglomerate Roche Holdings in advance of the Stock Exchange announcement. Following the announcement, stocks rose by more than \$8 a share.

In a written statement, Mr Raab said that his wife did not profit from the stock trades. "She knows it was wrong to have told her brother, who was having prolonged financial difficulty", Raab said, "and she accepts her penalty." He added, "we are both sorry for any embarrassment this incident has caused to Genentech and its employees".

Diane Gershon

FRENCH EDUCATION

Students hit jackpot

Paris

AFTER a week of nationwide protest marches and sit-ins, French *lycée* students have won significant new concessions from the education minister, Lionel Jospin (see *Nature* 348, 271; 22 November 1990). Complaining of dilapidated buildings, shortage of teachers and classroom violence, the students — with the backing of President François Mitterrand — have not only persuaded the government to

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

French students: tower of strength.

plough an extra FF4,000 million (\$800 million) into the 1991 education budget but also to bring in immediate reforms in the way secondary schools are run.

These complement existing ambitious and progressive reforms, involving greater decentralization and teacher recruitment, set in place on the advice of an expert committee which reported last year.

The highest priority in Jospin's latest reforms is the renovation of buildings, adding more study rooms, recreation facilities and libraries. An extra FF100 million (\$20 million) will be earmarked for better equipment. Second, students will be given a voice in decision-making and rights to a school press, to organize meetings and to distribute posters. Third, class sizes will be limited to 25 or 30 according to whether the school is vocational, technical or general. Classes of up to 40 have been reported. Fourth, student representatives will be consulted on proposed changes in teaching practice. Finally, an extra FF200 million (\$40 million) will be divided between *lycées* to provide a direct social aid fund for poorer students.

Jospin has also appointed a 'national correspondent' to report on progress of the national 'emergency plan' and expects the first reports at the end of November. The feeling in Paris is that, for once, a student revolt has been defused without any government casualties and with an intensification rather than a reversal of policy.

Peter Coles

Freedom for council

Sydney

THE National Health and Medical Research Council (NH&MRC), Australia's leading federal organization for funding medical research, has won some freedom from government control following the recent introduction of legislation that will make it a statutory body with its own budget. But despite its newly won independence, the council is coming under increasing political pressure to change its ways. At its national congress earlier this month, the council was pressed by both politicians and consumer organizations to pay more attention to social issues and to reduce emphasis on biomedical research.

Since 1938, the council has operated under an order-in-council from the executive of the British monarch. Under such an order, the government could change or abolish the council without debate. "The possibility no longer exists that the council could be dismissed at the whim of the government of the day", says chairwoman Diana Horvath.

Tania Ewing

The meaning of Wittgenstein

SIR—John C. Marshall's resentful rambles¹ on the significance of the philosophical heritage of Ludwig Wittgenstein provide little useful information for the uninitiated. The image of Wittgenstein, portrayed under the pretext of a book review, is that of a misguided and tortured showman whose main talent seemed to consist in exploiting his considerable intellect in the corruption of innocent minds with an error-ridden, intoxicating brew of mysticism and linguistic trickery. A few remarks may, perhaps, help towards a more balanced view of Wittgenstein's work than that allowed by Marshall's sadly vacuous 'hatchet job'.

Wittgenstein's later philosophy informs much debate at the interface between the philosophy of mind/language and the special sciences. In particular, the conclusions of his considerations on the nature of rule-following and the normativity of meaning are of some concern in much psychosemantic theory (the discipline concerned with the meaning or 'content' of psychological states), especially the area of biosemantics², and it is a contemporary question whether his insights can be made to cohere with the basic tenets of chomskyan theoretical linguistics³.

In addition, the scientific pertinence of Wittgenstein's writings is also exhibited in the wittgensteinian tenor of Donald Davidson's discussion of the relation between the mental and the physical, in particular the nomological irreducibility of the former to the latter⁴. Finally, Wittgenstein's remarks on the impossibility of a "private language" constitute a powerful critique of the tendency to offer objectifying accounts of the subjectivity of psychological states, perhaps best exemplified by the uncritical willingness of many scientists to model the cognitive space of the subject, the subject's 'inner life', on the literally inner goings-on of a computer.

Of course, that the brain, as a natural computational mechanism, is the *de facto* 'seat of consciousness' is undeniable, only the relationship between the two remains stubbornly opaque, except, possibly, for Marshall himself. This is no surprise, however: after all, it was Wittgenstein who said: "someone unpractised in philosophy passes by all the spots where difficulties are hidden in the grass, whereas someone who has had practice will pause and sense that there is a difficulty close by even though he cannot see it yet"⁵.

A. J. GREENFIELD

Department of Biochemistry and
Molecular Genetics,
St Mary's Hospital Medical School,
London W2 1PG, UK

MARSHALL REPLIES—First, "modern philosophers — like Wittgenstein — have no

theory of thought to speak of"⁶. Second, "in the specific case of Quine and Wittgenstein, ... the restrictions that they impose simply exclude from serious study the many fascinating questions they themselves raise"⁷. Third, in the absence of "any definite conception of body"⁸, the question of the irreducibility of the mental to the physical cannot even be formulated. Fourth, LISP is no more (and no less) a 'private language' than is English⁹. Fifth, certain aspects of 'consciousness' can be studied within the framework of natural science^{10,11}. But as Wittgenstein himself wrote: "nothing seems to me less likely than that a scientist or mathematician who reads me should be seriously influenced in the way he works"⁵.

The glory of intellectual and cultural life in the Austro-Hungarian Empire had an evil 'shadow-side' — apocalyptic kitsch — to which Wittgenstein rapidly succumbed and from which he never freed himself. In 1946, he is recommending dropping a few more atom bombs to make an end of "our disgusting soapy water science"⁵. I rest my case.

JOHN C. MARSHALL

Neuropsychology Unit,
University Department of
Clinical Neurology,
The Radcliffe Infirmary,
Oxford OX2 6HE, UK

1. Marshall, J.C. *Nature* **347**, 435 (1990).
2. Millikan, R.G. *Language, Thought and Other Biological Categories* (MIT Press, 1984).
3. Wright, C. in *Reflections on Chomsky* (ed. George, A.) (Blackwell, Oxford, 1989).
4. Davidson, D. in *Essays on Actions and Events* (Oxford University Press, 1986).
5. Wittgenstein, L. *Culture and Value* (Blackwell, Oxford, 1980).
6. Fodor, J.A. *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (MIT, Cambridge, 1987).
7. Chomsky, N. in *Philosophy, Science and Method* (eds Morgenbesser, S., Suppes, P. & White, M.) 260–285 (St Martin's, New York, 1969).
8. Chomsky, N. *Language and Problems of Knowledge* (MIT, Cambridge, 1988).
9. Fodor, J.A. *The Language of Thought* (Crowell, New York, 1975).
10. Marshall, J.C. & Halligan, P.W. *Nature* **336**, 766–767 (1988).
11. Weiskrantz, L. *Proc. R. Soc. B* **239**, 247–278 (1990).

Nowhere for the waste?

SIR—Our striving for a riskless society, coupled with political manoeuvring and a misinformed public, can lead us to heights of absurdity that would be funny were it not for the serious consequences in terms of cost, loss of productivity and misdirected effort. The US policy on the management of low-level radioactive waste (LLRW) is a case in point, with which every researcher who uses radioisotopes ought to be concerned.

Most US users of radioisotopes do not know yet that, unless they happen to live in the northwest or California, radioactive

waste generated in their laboratories will have no place to go after 1 January 1993, except perhaps to some interim storage facility on or off-site. This is because the existing waste sites will be required by law to close their doors to non-compact state generators, and other new regional sites have not been built. Congress mandated 11 regional waste sites (compacts) in 1980. Since then, the volume of LLRW has been reduced by 50 per cent. It would make sense to revise the legislation, and reduce the number of sites by consolidating the regions, but few in Congress wish to raise such a controversial issue.

Implementing the "below regulatory concern" (BRC) regulations issued by the NRC would also reduce waste volume by another 30 per cent. In that case, one or two sites could easily handle the LLRW for the entire country. But no sooner had the NRC announced the BRC regulations than Congress began considering legislation to rescind it. BRC regulations are based on sound, technical grounds, but our politicians seem to know better. They know that 1–10 mrem per year of additional dose per person as a result of BRC will endanger our health. But they do not understand that a round-trip coast-to-coast flight will result in similar dose rates or that this additional dose needs to be compared to the 360 mrem per year on the average that each of us is exposed to. They also overlook the fact that the dose rates were derived using extremely conservative models or that the impact of indefinite on-site storage on personnel exposures could be much higher.

LASZLO BERES

17 Galleys Hill Road Extension,
Hull, Massachusetts 02045,
USA

Specie not species

SIR—It is indeed gratifying to read that the nomenclature proposed for the earliest-known reptile is *Westlothiana curryi* (*Nature* **346**, 399; 1990) and that Wood's name will not be cited in the nomenclature.

I and many others deplore the mercenary aspect that is becoming increasingly prevalent and, as a palaeontologist, I am proud to say that in nearly 50 years of scientific work I have never sold a single fossil — all have been donated to museums and institutions solely in the interests of science.

Nevertheless, the greed that led to a 'value' of £205,000 being suggested for the reptile should be commemorated in some way. Perhaps a new suborder should be erected for *Westlothiana curryi* — might I suggest *Avariceoidea*?

P. J. SMART

46 Brecon Way,
Bedford MK41 8DD, UK

Cooking in the sunshine

Daniel M. Kammen and William F. Lankford

'Appropriate technology' is touted as a solution to many of the energy needs of developing countries. A pilot project to introduce solar ovens in Central America shows one way forward.

"EVERY year the forest moves further and further from my house; now I have to walk three hours a day to collect what wood I can." The speaker is an Indian woman from the Pacific plain of Guatemala, where clearing for export farming has all but eliminated the forest sources of fuel-wood. Two thousand million people worldwide depend primarily on wood for cooking, and half of the annual world wood harvest of three thousand million tons is used as fuel. In developing countries, over 50 per cent of all energy is used to prepare food, which in turn consumes two-fifths of the already scarce capital. By contrast, developed countries expend less than 5 per cent of their total energy consumption on cooking¹.

A solution to this unhappy situation must involve a combination of conservation and increased reliance on alternative (ideally renewable) energy technologies. Solar cooking is a natural avenue to pursue because many developing countries are equatorial and receive over 1,100 W m⁻² of solar radiation at normal incidence (mid-day) and up to 6,300 W m⁻² over the course of a day². The history of attempts to introduce solar cooking on a large scale, however, is poor. In the 1960s and 1970s ovens that concentrated solar energy using parabolic dishes, and simple box-cookers, were tested among American Indians in Arizona, in villages in northern Mexico and in the Middle East. In our view, these projects failed to gain acceptance because they were approached as donations of technology rather than as collaborations.

Over the past three years we have been working with a variety of community groups in Guatemala, El Salvador, Honduras, Nicaragua and Costa Rica to introduce a simple box-type solar cooker (described overleaf). In Guanacaste province, Costa Rica, a group of nine women who worked together on a radio talk show concerned with women's rights worked with us to organize a demonstration of solar cooking. The Women's Solar Committee of Oriente, a village of 80 families, was formed. Each woman built an oven, and the group constructed three more ovens to be used in promoting future projects.

The protocol that has emerged from our experience, and is now used in each workshop, consists of five stages. Initial planning is completed under the leadership of a respected local person, and an oven is

given to a member of the community for a trial period. The oven's capabilities are publicly demonstrated and the theory of oven performance is discussed while a 'solar' meal is prepared. A local group, formed from the audience at the demonstration, then completes an oven workshop in which materials, tools and instruction are provided; the participants provide the labour and each receives an oven. A follow-up team is then formed to make regular house visits to monitor oven use and solve problems that may arise. Finally, the workshop group obtains legal status and financial autonomy through domestic or international grants, and further collective projects are initiated.

The local community itself carries out a screening process to select the participants, who all sign an agreement to take part in future workshops and to return their oven if it is not used regularly. That second clause has never been invoked — the women in Oriente so like the ovens that they regularly gather for solar *banquetes* where recipes and techniques are exchanged. In the summer of 1990 a second workshop was held with the Solar Committee serving as instructors for others in the community.

The final stage of the protocol encourages further box-cooker workshops as well as additional community initiatives in technology — for example the construction of high-efficiency communal wood stoves, wells, solar-powered refrigerators or larger ovens that permit shared use^{3,4}. Projects with the potential to generate much-needed capital include high-diversity (kandian) garden plots; reforestation schemes; the acquisition of a community fishing boat; or rattan or latex plantations. Bolstered by the oven project, the Oriente group now plans a communal garden, which has attracted government interest and support.

During the rainy season, cloud cover may prevent a meal from completely cooking in the solar-box. Literally as insurance against a rainy day, the women of Guanacaste adopted a high-efficiency wood-burning stove, and the Solar Committee has begun making and selling these stoves with great initial success. The history of the Oriente group thus exemplifies the keys we found to high oven use: that the individuals selected make a considerable investment of their own time and energy in the project, and that they work together as a group. ►



Spiritual approval — a parish priest in Guatemala blesses the solar ovens with holy water.

Many previous oven designs involved several collecting mirrors or telescopic focusing of sunlight to cook food by radiative heating^{3,4}. Although ovens such as these reach higher temperatures (up to 250 °C) than the box-type model, and can in principle be operated by any member of the community, they are inconvenient to use. Focusing ovens require constant adjustment to direct the sun onto the cooking vessel, are generally unstable in wind and cannot be constructed locally. By contrast, the box-cookers can be fabricated, or if necessary repaired, with materials and carpentry techniques that are familiar worldwide. In addition, parabolic solar concentrators are not generally designed to take more than one cooking vessel. These drawbacks, and the need for the cook to stand in the hot sun over a blinding reflector, in part explain the failure of previous attempts to introduce solar cooking in developing countries.

The oven design with which we began the Central American project has undergone steady adaptation⁵. At each new workshop the participants suggested refinements, such as deepening the cooking space to accommodate a variety of pot sizes and adding a support table mounted on wheels. As the project progressed, the oven was tailored to fit local needs.

To study the effect of these changes, each of the seven main workshop sites was revisited and evaluated by one of us (W.F.L.) during the late summer of 1990. The ovens constructed in August 1988 at the earliest workshop in La Trinidad, Nicaragua, are almost never used (an average of 0.0 to 0.4 oven uses per week). In contrast, those constructed at the end of the project, in Guapinol and Sololá, Guatemala, however, are in constant use, from 5.2 to 6.5 days a week.

The project has come full-circle; the ovens empower the workshop participants who in turn adapt the technology, integrating it into the resource base of the community. The failure of numerous well-intentioned appropriate technology projects and, in particular previous attempts to introduce solar ovens in the developing world, can be traced to the imposition of unfamiliar technology without the social interplay and development of self-reliance.

In the oven project, we found that women are much more capable participants than the typical Latino man — their patience, dedication to precision, receptivity to learning new skills and lack of *machismo*, as well as their traditional role as family cook, all contribute to this. Women perform up to 80 per cent of the subsistence agricultural work and 90 per cent of domestic tasks in Africa⁶; the percentages are somewhat less skewed in Central America. It was women, both as participants and as group leaders, who proved critical to the success of the oven

The solar box-cooker

The simple glass-topped solar box-cooker, first popularized in the 1950s^{2,5,7}, remains the basic design today and is the prototype for our work in Central America. The glass transmits visible light, but is opaque to re-radiated infrared wavelengths: the ovens are miniature greenhouses. The pots containing the food rest directly on a conducting metal floor plate (both painted black) so the oven does not need to be constantly moved to face the sun. The ovens can reach temperatures of 150 °C in less than an hour, and can exceed 170 °C. The lid is covered in aluminium and doubles as an extra solar collecting panel. The ovens have been used to boil, bake, simmer, braise and sauté foods, and to pasteurize naturally contaminated water⁸; in moderate to strong sunlight a full meal can be cooked in 2–4 hours.

project. This fact should act as an impetus to multinational development agencies such as the World Bank, Inter-American Development Bank, and UN sponsored programmes (UNDP) and USAID, to emphasize educational, agricultural and business opportunities directed specifically to women. The World Bank recently launched several long-overdue educational and economic opportunity programmes for women that we hope will serve as models for other such initiatives.

But what about the economics of the oven project? On average, a poor Central American family can contribute only 25–50 per cent of the cost of the construction materials, which range from US\$80 to \$100 for each oven. This represents two to three months wages in local currency, and results from the need to use materials (glass, reflecting foil, metal sheeting and household hardware) often priced on the international market. The cost of an oven in terms of fuelwood purchases averages 6–12 months' wages, and is relatively constant throughout the region. An Indian woman from the Guatemalan highlands brought home to us the strain of the never-ending search for firewood. The woman, dressed in an intricately woven *traje*, or traditional dress, thanked us for saving her weaving. It turned out that previously most of her working day had been spent in collecting wood, but with a solar oven in the house she now had time to spend at her loom.

There are a number of technical and cultural lessons to be learned from our experiences in attempting to introduce appropriate technology in the developing world. But it is essential to keep in mind that solar ovens are only one component of a solution that must integrate such disparate aspects as basic conservation, economic empowerment of the poor, diversification of energy sources and reforestation — technology is only a tool, not a solution.

The introduction of 'appropriate tech-

nologies' on a large scale must be approached with a small-project mentality: the technology and the protocol must be tailored to local needs, with careful follow-up. As a general rule, developing nations would be better aided if large multinational agencies increased their level of support for promising small-scale projects. Helping non-governmental organizations and direct funding of proposals originating in local communities are two obvious avenues to pursue.

The beneficial effects of the solar oven project can be seen on many levels. Cooking fires, frequently located in confined indoor spaces, are consistently cited as one of the main sources of respiratory, skin and eye infections and ailments among women and children in developing countries; indeed, the prospect of smoke-free cooking was often the reason why a group requested a solar workshop. And even a modest number of solar ovens can have an important impact on the local environment. We have worked primarily with the rural poor, who are often relegated to economically marginal land. These areas, typically on the edges of swamps, deserts and rain forests, are also the most fragile ecosystems that can least support extraction of fuelwood.

At a regional level, the need for fuel leads to harvest rates that 'mine' the forests of their timber capital at unsustainable levels. The end result of this cycle can be seen in Africa, where in some countries only 10 per cent of the demand for fuelwood is satisfied. Dung is then diverted from use as fertilizer to use as fuel. The size of the problem cannot be overstated — the 400 million tons of dung burned each year in Africa and Asia represent enough fertilizer to grow 20 million tons of grain, enough to feed 100 million people for a year¹.

International efforts at conservation could also be augmented by encouraging local governments to subsidize oven construction, as is the case in India. Given the enthusiastic reception that the solar ovens received in Central America, we believe that the project, properly adapted, could work anywhere. □

Daniel M. Kammen is in the Division of Biology, Mail Stop 216-76, California Institute of Technology, Pasadena, California 91125, USA. William F. Lankford is in the Department of Physics, George Mason University, Fairfax, Virginia 22030, USA.

1. Myers, N. *The Primary Source. Tropical Forests and Our Future* (Norton, New York, 1984).
2. Telkes, M. *Solar Energy* **3**, 1–11 (1959).
3. Garg, H.P., Mann, H.S. & Thanvi, K.P. in *Sun. Mankind's Future Source of Energy* **3**, 1491–1496 (1978).
4. Tabor, H. *Solar Energy* **10**, 153–157 (1966).
5. Kammen, D.M. & Lankford, W.F. *Solar Wind Technol.* **7**, 463–472 (1990).
6. Warford, J.J. in *Environmental Management and Economic Development* (eds Schramm, G. & Warford, J.J.) 7–23 (Johns Hopkins Univ. Press, Baltimore, 1989).
7. Nandwani, S.S. *Solar Wind Technol.* **5**, 159–170 (1988).
8. Ciochetti, D. & Metcalf, R.H. *App. env. Microbiol.* **47**, 223–228 (1984).

Cluster arithmetic made simpler

Clusters of small numbers of identical atoms are all the rage, but the calculation of their properties taxes even large computers. That is why a simpler method deserves a welcome, but it also has heuristic value.

THE cluster business, the study of small numbers of identical atoms held together by a combination of chemical bonds and electrostatic forces, is booming away, and for a variety of reasons. As always, new techniques have had a powerful influence, notably those allowing for the self-assembly of clusters from the ingredients of well-cooled atomic beams. A further incentive is the prospect that atomic clusters may have important practical advantages, as catalysts for example, or possibly as sub-components of electronic devices. Evidently there are prizes to be won.

But the cluster business has also turned up surprises, not least that the configurations of small clusters are often strikingly different from what would be expected from the interatomic configurations in the bulk solids. Both the four-atom and six-atom clusters of lithium are planar structures; in Li_4 , the atoms are at the vertices of a rhombus, in Li_6 , three atoms are at the vertices of an equilateral triangle and the others at the mid-points of the three sides. Why should these unexpected structures be stable? And, more generally, what can be said about the transition from small clusters to the interatomic configuration of bulk solids?

These are questions from which Osamu Sugino and Hiroshi Kamimura of the University of Tokyo embark on a simplified way of calculating the stability of small clusters (*Phys. Rev. Lett* **65** 2,696; 1990). Their starting-point is a set of data for the stability of small lithium-atom clusters. In the process, they describe a novel kind of bond between atoms, a "glue bond" as they call it.

That the cluster business needs a simplified method of calculation is not disputed. The techniques of computational chemistry may be used to calculate the properties of an arbitrary atom cluster, which for these purposes is simply another kind of molecule. For lithium clusters, for example, in which each atom is a singly charged helium-like core with a single external electron, it would be reasonable to begin with the assumption that there is one potentially bonding electron on each atom. For Li_4 , it would then suffice to calculate the energy of the four-electron system for all possible configurations of the four nuclear cores.

The snag, as Sugino and Kamimura note, is that even those with access to supercomputers are lucky to win the time

required for comprehensive calculations. For each cluster is intrinsically more difficult than a molecule. In principle, the configuration and even its symmetry may not be known, while telling what is the optimum configuration depends on a comparison between that and all the others possible. The calculation of clusters with large numbers of atoms is, for practical purposes and at least for the time being, only exceptionally feasible.

So is there a simpler way? Sugino and Kamimura go back to an earlier attempt at generalization by Mark H. McAdon and William A. Goddard (*Phys. Rev. Lett* **55**, 2,563; 1985), also based on calculations of lithium clusters. Their objective was to construct configurational rules on the basis of accurate calculations of a few simple lithium clusters in a few simple configurations.

One of their more striking findings was that the distribution of electron density in, say, a square Li_4 configuration is concentrated about the sides of the square, or symmetrically about the mid-points of the interatomic bonds to be precise, and not at the atoms themselves. Those who choose to regard this as a hint of the electron delocalization found in the conduction band of an electronic solid, a metal for example, will not be mistaken.

McAdon and Goddard also found that each of the electron orbitals in Li_4 is occupied by a single electron only, with the spins of neighbouring electrons paired (in the sense of being antiparallel) in pairs. The distortion of the square into a rhombus is advantageous because of the physical repulsion between parallel (or Pauli-excluded) electrons. Another way of putting this is to say that the acute angle of the rhombus (measured at about 60°) includes two paired electrons in a roughly spherical orbital filling that triangular half of the rhombus.

What Sugino and Kamimura have now done, in the hope of simplifying accurate calculations of lithium clusters, is to embark on quantum chemistry with electron orbitals which more accurately represent this kind of electron distribution. They represent them as ellipsoids centred on moveable points and, allowing for the electrostatic repulsion between the helium-like cores of the lithium atoms, carry through a calculation to determine not only the parameters defining the ellipsoids but the positions of the nuclei and the points around which the different elec-

trons are centred. These are what they call their glue-bonds.

The outcome of the calculations is impressive. First, the stability (or the negative energy) of lithium clusters is accurately reproduced up to the limit of strict *ab initio* calculations based on self-consistent field optimization. Li_8 , for example, emerges (as it should) as one of the most stable clusters. It is a curious structure, which may be considered as two Li_4 rhombi, one above the other, which are puckered into boat shapes in opposite directions with one turned through 90° . There are four glue-bonds, each occupied by a pair of electrons, which are themselves arranged tetrahedrally (whence, say the authors, the stability).

The authors carry their calculations beyond Li_{14} , the most complicated cluster to yield to self-consistent field treatment, as far as Li_{36} . They conclude that there are "magic numbers" of stability at clusters of 8, 14, 18, 20, 26, 30 and 34 atoms, although, with almost half the even numbers between 8 and 36 in the list, there may be too much magic for some tastes. The numbers 8, 20 and 30 might be a more prudent choice, to judge from the results of the calculations.

What does this mean for the transition from cluster to bulk solid? The optimum configuration for any cluster is evidently a trade-off between the electrostatic energy and the electron exchange or Pauli-exclusion energy. The case of Li_4 illustrates that an increase of the former on distortion of a symmetrical structure may be more than offset by a decrease of the latter involving the physical separation of glue-bonds from each other.

What Sugino and Kamimura argue is that there comes a point, as clusters become larger, at which it is no longer energetically favourable to place all the atoms on the outside surface of an open network. Then, there is less scope for geometrical distortion of the kind required if exchange energy is to be minimized, so that magic numbers melt more completely into the background.

The calculations show that this transition to bulk constitution must be under way at Li_{26} , the least energetic form of which has glue-bonds in a pattern with icosahedral symmetry, which in turn implies that one of them is enclosed within a shell of all the others. The cluster is not yet a metal, but nearly so.

John Maddox

Why the long latent period?

Charles R. M. Bangham and Andrew J. McMichael

THE central problems in understanding the progression of infection with human immunodeficiency virus (HIV) to AIDS stem from the virus's long latent period: the median time from infection to development of the disease is about eight years. Two questions then arise. First, what is the role of the immune response in the progression to AIDS? More precisely, is the decline in the immune response the cause or consequence of the rise in HIV titre, and do antibodies or cytotoxic T lymphocytes (CTL) exert significant selection against HIV variant sequences? Second, what is the effect of the extreme sequence diversity of HIV: do variants arise that escape the immune response¹? Nowak and colleagues, in a report published earlier this month in the journal *AIDS*², use a simple model to try to explain the long latent period on the basis of the sequence diversity of HIV. According to the model, AIDS is the result of an increase in the number of immunologically distinct variants of HIV above a critical threshold. How can this hypothesis be tested?

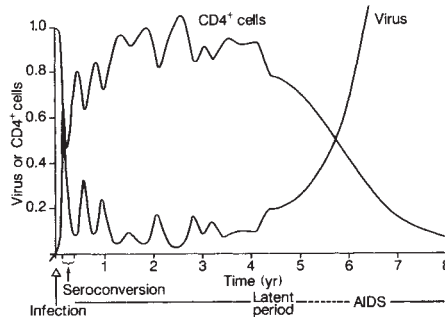
RNA viruses, such as HIV, are highly variable in sequence³, largely because of the absence of a 'proof-reading' 3'-5' exonuclease activity in the RNA-dependent polymerase. Variant sequences of an RNA virus arise frequently both *in vitro* and *in vivo*; the diversity within one infection, particularly a long-term persistent infection, may be greater than the diversity between infectious isolates, because many of the variants produced are defective in replication or infectivity, and so are not propagated. The mixed population of variant sequences of a virus has been called a 'quasispecies'⁴.

The main assumptions made in the new model are these: (1) HIV can kill any CD4⁺ cell, but only a subset of CD4⁺ cells are involved in killing a given variant of HIV; (2) antigenic variants of HIV arise which escape destruction by the immune system. The basic equation in the model is then $dV_i/dt = V_i(r-sz-px_i)$ where V_i is the population size of the variant sequence i ; t is time; r is the net replication rate of the virus (initially assumed to be the same for each strain); z is the number of 'immune agents' (antibodies or T cells) that recognize conserved epitopes; x_i is the number of immune agents specific to the variant sequence i ; and s and p are rate constants respectively for the effect of the cross-reactive and specific elements of the immune response.

On this model, the immune system can be said to control the variant i if $dV_i/dt < 0$, that is if $r-sz-px_i < 0$. Summing over all n variant sequences of the virus, we have

$n(r-sz)-p\sum x_i < 0$, that is $n < p\sum x_i/(r-sz)$. The authors show that when the number n of HIV variants exceeds a critical value $n_c = p\sum x_i/(r-sz)$, the total virus load increases steadily with time.

The model includes a stochastic term which simulates the emergence of variants that are initially not neutralized by the pre-existing immune response. As the immune system responds, each successive variant is temporarily controlled, until the number of variants exceeds the threshold



Changes in the numbers of CD4⁺ cells and HIV titre (arbitrary scale), following infection with HIV, predicted by the model proposed by Nowak *et al.*².

value n_c , after which all the variants can multiply unhindered and kill the remaining CD4⁺ cells.

The authors then show that a more realistic model, including notably the assumption that different HIV variants replicate at different rates, does not qualitatively alter the outcome. The resulting changes in the CD4⁺ cell number and HIV titre occurring over the years during progression to AIDS are summarized in the figure, and they correspond well to what has been observed in humans⁵.

Attempts to explain the initial control of HIV and the eventual breakdown of the immune system in terms of simple immunological mechanisms are fraught with danger because the population dynamics of cells and virus rapidly become complex, and this complexity is increased greatly by the sequence diversity of the virus. The usefulness of models such as that proposed by Nowak *et al.* is to identify what seem to be the important factors determining disease progression. The clear implication of the model is that the most important single factor leading to the breakdown of immune control of HIV is the increase in the diversity of the virus. If this is correct, we should not be examining the predominant HIV sequence(s) isolated during the symptomatic phase of HIV infection to look for 'determinants of virulence'. This variant may simply be a fast-replicating one that was around when the immune system broke down.

However, tests of the hypothesis will

have to be formulated with great care. The principal difficulty is that what we need to measure is not the sequence diversity but the antigenic diversity of HIV: the assays should therefore be immunological, and should not rely simply on sequence comparisons. Further, detection of variant HIV sequences depends mostly on the use of the polymerase chain reaction (PCR), but this technique fails to distinguish between functional and defective variants of HIV. We know that a high proportion of integrated proviruses are defective, 10–15 per cent in *tat* alone⁶. Presumably we should therefore try to identify replication-competent variants by growing them in cell culture: the snag there is that HIV continues to undergo mutation *in vitro*⁶. These problems are magnified in the silent phase of HIV infection, during which the viral abundance is low: it is therefore still more difficult to get an accurate measure of the number of variant sequences, and to culture the virus for long periods is to select for fast-replicating variants.

A logical approach is perhaps as follows: prepare complementary DNA from the peripheral blood lymphocytes of an HIV-infected person either directly or after a brief period (say, up to one week) in culture, so that replication-competent strains begin to grow but do not diversify further (and one fast-replicator does not predominate); then amplify by PCR the regions of the HIV genomes that are known to encode important CTL, or antibody epitopes, such as the V3 loop of Env (antibody epitope) or the Gag (CTL epitopes) proteins. There is variation in the sequence of Gag CTL epitopes, even in highly conserved regions of *gag* (ref. 7; R. E. Phillips, personal communication). This approach involves not only sequence determination, but both antibody and CTL assays to test the antigenic distinctness of a new variant. And it depends on a knowledge of the most important epitopes in subjects of known HLA genotype.

A potentially important corollary of the model, noted by the authors, is that any agent such as a drug which reduces the replication rate of the virus should increase the critical number of HIV strains required to cause AIDS, and so reduce the chance of disease progression. Is this how AZT works?

Charles R. M. Bangham and Andrew J. McMichael are in the Institute for Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK.

1. Pircher, H. *et al.* *Nature* **346**, 629–633 (1990).
2. Nowak, M. A., May, R. M. & Anderson, R. M. *AIDS* **4**, 1095–1103 (1990).
3. Steinhauer, D. A. & Holland, J. J. *A. Rev. Microbiol.* **41**, 409–433 (1986).
4. Eigen, M. *Naturwissenschaften* **58**, 465–523 (1971).
5. Miedema, F., Tersmette, M. & van Lier, R. A. W. *Immun. Today* **11**, 293–297 (1990).
6. Meyerhans, A. *et al.* *Cell* **58**, 901–910 (1989).
7. Goodenow, M. *et al.* *J. Acq. Immun. Def. Syn.* **2**, 344–352 (1989).

Nanostructured materials

Robert W. Cahn

WHEN a new field of research arrives upon the public scene, the first thing apt to excite dissension is either the naming of the key phenomenon, or else its very existence. Following the recent claims concerning fusion of deuterium dissolved in metallic palladium, the term 'cold fusion' was happily accepted by all, but the phenomenon was not. On the other hand, although the theme of a recent conference entitled *Materials with Ultrafine Microstructures** was agreed as being sound enough by everyone present, the proper nomenclature was not.

The subject of the conference, the first full-scale meeting on this theme, was the synthesis, characterization and properties of polycrystalline materials, the constituent grains of which are only a few nanometres across (typically, $2\text{--}25 \times 10^{-9}$ m). The originator of this field research, H. Gleiter (University of Saarbrücken) coined the term 'nanocrystalline materials'. Because such materials are usually made by consolidating minute clusters of atoms, the term 'cluster-assembled materials' was the next to be proposed¹. As the title of the recent conference indicates, others like the sound of 'materials with ultrafine microstructure', while a further, slightly eccentric group proposes the term 'finite systems' to embrace clusters and cluster assemblies — and quasicrystals for good measure. But 'nanostructural materials', the title of a projected international meeting in 1992, may be the finally accepted description.

Clusters

The main emphasis of the meeting was on synthesis, processing and applications. Until recently, Gleiter's original method of making the precursor clusters, by evaporating a metal or ceramic into a noble gas at fairly high pressure, restricted quantities to fractions of a gram. Krätschmer and colleagues recently showed² how buckminsterfullerenes such as the spherical cluster C_{60} could be made in gram quantities by heating a carbon rod electrically and condensing the resulting smoke. (C_{60} has captured popular imagination to such a degree that a cartoon recently had the hero thrashing about helplessly in a vat of low-friction C_{60} fullerene.)

Clearly, the urge to scale up is now widespread. For example, the straight evaporation method has been developed to the point where useful nanostructured metallic films can be deposited on a silicon substrate and aerosol clusters can be directed to form conducting strips no

more than 50- μm wide at specific locations on microcircuits (M. Oda, Vacuum Metallurgical Company). These films and wires have good adhesion and conduct well. Oda claimed that his method had distinct advantages over conventional methods of forming 'wiring' on microcircuits. The size of circuit elements made in this way is being steadily reduced, and (this being Japan) the equipment needed is already for sale. M. Uda (Nisshin Steel Company, Japan) showed that the rate of generating metallic clusters into hydrogen gas, in preference to argon, could be enhanced by several orders of magnitude by first activating the hydrogen electrically. The lone Soviet participant, L. Trusov (Research Industrial Enterprise Ultram, Moscow), though he was coy about his methods of synthesis, showed favoured participants a large metallic nanofilter sheet made from evaporated and consolidated clusters.

Combination

Probably the most impressive account of synthesis and application combined came from B. Kear's group (L. McCandlish, Rutgers University). They set out to improve the 50-year-old 'hard metal' which consists of rather coarse, hard and brittle tungsten carbide crystals dispersed in a soft, ductile matrix of cobalt metal — a cermet. They guessed that if they could make such a material with nano-sized crystallites, its mechanical properties might be enhanced. Enlisting the aid of a chemist, the group combined *tris*-(ethylenediamine) cobalt tungstate with ammonium metatungstate and cobalt chloride, carburized the product first with CO and then with a CO/CO₂ mixture (a sequence which permitted them to combine rigorous control of C activity with rapid production), and then consolidated the mixed product to form nanostructured hard metal. The new material is up to 40 per cent harder than conventional hard metal for a given cobalt content and the friction and wear rates are much lower; high hardness and low wear rate is the required combination for machining tools. An industrial-scale plant to make the new material has now been designed around the use of a fluid-bed reactor.

Also from Rutgers University, S. Danforth analysed in detail the benefits flowing from the use of nanostructured silicon nitride (one of the front-running candidates for ceramic automotive engine components). Essentially, such a fine powder can be sintered at much lower temperature than conventional powders and this permits the product's porosity to be depressed to low levels and its con-

sequent properties, especially at very high temperatures, to be improved. The ready sinterability of nanostructured powders formed the theme of several other papers and is clearly an important research topic: it is important that grain growth is often surprisingly sluggish so that nanosized powders can be sintered without coarsening much. The reasons for the slow growth constitute another current research theme.

Strength

The strength of nanocrystalline Cu and Pd seems hardly to depend on grain size d : the familiar Hall-Petch $d^{-1/2}$ dependence was generally not observed (J. Weertman, Northwestern University). A particularly novel study (V. Provenzano, Naval Research Laboratory, Washington D.C.) building on much earlier studies by Marcus³ (experimental) and Louat⁴ (theoretical) shows that materials of quite exceptional strength — they are in fact termed as 'superstrong' — can be obtained by embedding a high volume fraction of nanosized particles in a ductile matrix or alloy; in one version, the matrix is glassy. Apart from the ease of sintering nanosized powders, another advantage of consolidated nanostructured ceramics is that they can be superplastically deformed, as Gleiter first showed some years ago. This process, reported by several contributors and recently described in this journal⁵, depends on the diffusion of lattice vacancies through or round small grains, effectively transporting matter to respond to applied stress.

Not only people take an interest in nanostructures: nature does also. There are fascinating lessons to be learned from the fine structure of seashells, abalone shells (*halotis rubescens*) in particular (I. Aksay, Seattle University). One layer of such a shell consists of 250-nm laminations of aragonite (a polymorph of CaCO₃) separated by 20-nm layers of an organic compound. This structure generates both high strength and high fracture toughness. Years ago, useful lessons were learnt from the laminated fibre-reinforced structure of the wing-cases of rose chafer beetles;

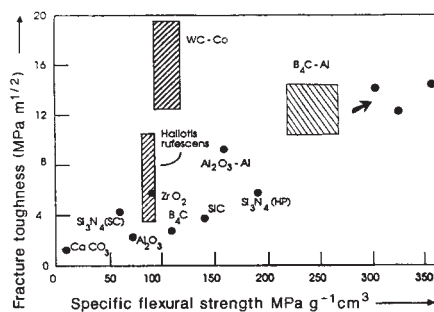


FIG. 1 Fracture toughness versus fracture strength (normalized with respect to density) of the nacre section of abalone shell compared with several high-technology ceramic and cermet materials. (After Sarikaya *et al.*⁶.)

* Acta Metallurgica conference on Materials with Ultrafine Microstructures Atlantic City, 1–5 October 1990.

RÉSUMÉ

To the point

ALTHOUGH conventionally regarded as point sources of charge, electrons may, according to some theories of high-energy physics, acquire an electric dipole — equivalent to a separation of charge. But new experiments have failed to detect any such dipole. The theories have been developed to explain a fundamental asymmetry — termed charge-parity violation — inherent in the weak nuclear force. Although different theories predict different values for the electron's dipole, it is expected to be vanishingly small. Using a 600-fold amplification of the dipole's value expected from relativistic effects in atomic tantalum, K. Abdullah *et al.* (*Phys. Rev. Lett.* **65**, 2342–2350; 1990) have sought, unsuccessfully, to find its imprint on this element's spectrum. The value the authors derive ($-2.78 \pm 8.3 \times 10^{-27}$ e cm (compare water's 3.9×10^{-31} e cm) rules out more recent models invoking the increasingly popular 'Higgs boson' to explain charge-parity violation. But a further improvement of ten orders of magnitude is needed in the accuracy before all mechanisms for this asymmetry are tested.

Quasi-stable

SOME quasicrystals may be more stable than periodic crystals, claims Z. Olami (*Phys. Rev. Lett.* **65**, 2559–2562; 1990). Quasicrystals have no long-range periodicity yet give sharp diffraction patterns with 'forbidden' symmetries (tenfold, for example, suggesting a quasiperiodic lattice of icosahedral symmetry). The first quasicrystals discovered, alloys of aluminium and manganese, were formed by rapid quenching, implying that they were being trapped in an uncomfortable, metastable configuration that would prefer to relax to a more stable, regular crystalline form. But by summing up interatomic interactions throughout the respective lattices, Olami finds that a monatomic face-centred icosahedral phase can be energetically more favourable than the most obvious periodic alternatives, such as body- and face-centred cubic arrays. So forming quasicrystals may prove easier than expected.

Face to face

FOLLOWING the positive identification of the skull of Wolfgang Amadeus Mozart last year, a group led by P.-F. Puech has used forensic techniques to recreate his face (*L'Information Dentaire* **22**, 2003–2008; 1990). The skull was exhumed in 1801, since when it has been kept in the Mozarteum in Salzburg. Reconstructing the face involves using clay to build up layers of soft tissue on a cast of the cranium. The nose is difficult to include because it depends on cartilage, but its dimensions can be calculated from the size and position of the nasal aperture — only the ears are impossible to reconstruct.

now⁶, a B₂C/Al laminar composite has been made, following the abalone model, and proves to have outstanding properties (Fig. 1).

By no means was the conference limited to applications. Fundamental scientific issues were also much aired: for example, the nature of the intercrystalline boundary regions in nanostructured materials. Whereas high-resolution electron microscopy (G. Thomas and R. Siegel, Argonne National Laboratory) and Raman spectroscopy of nanocrystalline CO₂ (ref. 7) indicate a normal, atomically narrow grain-boundary structure. Mössbauer studies (Gleiter; B. Fultz, California Institute of Technology), electrochemical measurements (R. Kirchheim, Max-Planck-Institut für Metallforschung, Stuttgart, Germany), density and diffusion measurements (Gleiter) and calorimetry indicate that boundaries are more disordered and have a lower local density than conventional boundaries. The curiously sluggish grain growth already cited may be linked with this. Fultz and coworkers' Mössbauer data for a nanostructured Cr-Fe alloy can be accurately analysed to yield an estimated boundary width of about 1 nm; after grain growth, the boundary signal of course diminishes. The issue is still very much open; it may be that the structure depends sensitively on the nature of the misorientation and varies from boundary to boundary.

Several excellent papers were devoted to the characteristic clusters themselves. S. Riley (Argonne National Laboratory) showed how precise measurements of NH₃ uptake could be used to identify the exact geometrical form of size-separated metal clusters, Ni and Co in particular, because NH₃ attaches itself only to atoms standing proud of the surface, for instance at icosahedral vertices. X. Miao and P. Marquis (Birmingham University) showed how pH control and the combination of two differently sized populations (of alumina and silica) allowed them to create stable combined colloid particles of the two compounds which then sinter more easily.

Magnetic studies of both elementary and multicomponent nanoclusters, as well as metallic multilayers, reveal curious combinations of ferromagnetic, anti-ferromagnetic and superparamagnetic behaviour, as well as anomalous easy magnetization directions. In this connection, a highlight of the conference was H. Warlimont's report on recent developments by G. Herzer in his laboratory (Vacuumschmelze, Hanau, Germany) of an innovation first reported by Yoshizawa *et al.*⁸. Certain metallic glasses containing Fe, Si and B have excellent soft magnetic properties and are currently used as transformer laminations. Yoshizawa and co-workers found that if Nb and (crucially) Cu are added to this glass and it is then

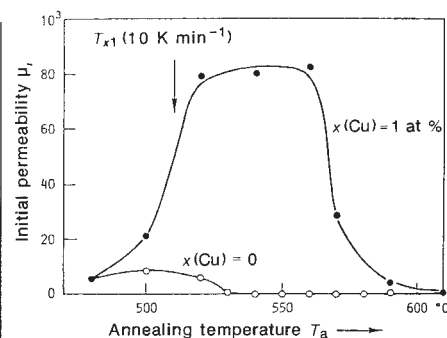


FIG. 2 Initial permeability, μ_i , of Fe_{74.5-x}Cu_xNb₃Si_{13.5}B₉ alloys, initially glassy (with $x=0$ and $x=1$), as a function of annealing temperature, T_a , for a constant 1-h annealing time. T_{x1} is the first crystallization temperature. (After G. Herzer and Warlimont.)

crystallized in the right way, a giant magnetic permeability can be attained, comparable with that of permalloy.

Herzer and Warlimont's results (Fig. 2) show that 1 atomic per cent of copper has a spectacular effect. It turns out that the annealed glass (provided it is not heated at too high a temperature) consists of nanocrystalline, ferromagnetic Fe₃Si phase separated by films of residual glass, rich in B and Cu, which is either nonmagnetic or has weaker magnetism. The residual glass crystallizes above 600 °C. Herzer and Warlimont analysed the magnetic properties of such a structure.

Nanocrystalline ferromagnetic particles are single-domain (Gleiter in a recent review⁹ cites very new evidence to this effect), and the ferromagnetic exchange interaction then overtrumps the effects of magnetocrystalline anisotropy in the nanograins and constrains the magnetization vector to be parallel to that in a neighbouring grain, irrespective of easy directions of magnetization. At a critical grain size, this results in very low coercivity and high permeability, as observed: here the critical dimension is around 35 nm. This Cu-bearing alloy has also a very low magnetostriction and the hysteresis loop shape can be shaped by magnetic annealing. This kind of nanostructured magnetic dispersion in a residual glass, by analogy with the 'superstrong' dispersions cited above, might perhaps be termed a 'super-magnet' — nanostructural researchers are predisposed to superlatives. □

Robert W. Cahn is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK.

- Andres, R.P. *et al.* *J. Mater. Res.* **4**, 704–736 (1989).
- Krätschmer, W., Lamb, L.D., Fostiropoulos, K. & Huffman, D.R. *Nature* **347**, 354–358 (1990).
- Marcus, M.A. *Acta metall.* **27**, 893–902 (1979).
- Louat, N.P. *Acta metall.* **33**, 59–69 (1985).
- Wakai, F. *et al.* *Nature* **344**, 421–423 (1990).
- Sarikaya, M., Gunnison, K.E., Yasrebi, M. & Aksay, I.A. *Mat. Res. Soc. Symp. Proc.* **174**, 109–116 (1990).
- Melendres, C. A. *et al.* *J. Mater. Res.* **4**, 1246–1250 (1990).
- Yoshizawa, Y., Oguma, S. & Yamauchi, K. *J. appl. Phys.* **64**, 6044 (1988).
- Gleiter, H. *Prog. Mater. Sci.* **33**, 1 (1990).

Clicking into decline?

John Maynard Smith and Sean Nee

TWENTY five years ago, H. J. Muller¹ pointed out that a finite asexual population may decline in fitness through the chance accumulation of deleterious mutations — a process known as Muller's ratchet². The process has been invoked to explain a variety of phenomena, ranging from clonal senescence in ciliates to the genetic deterioration of the Y chromosome. It is difficult, however, to produce a convincing experimental demonstration of the ratchet. In a report on page 454 of this issue³, Chao now presents evidence for its operation in an RNA virus.

The virus concerned is $\phi 6$, an RNA phage infecting the bacterium *Pseudomonas phaseolica*. Its genome consists of three segments contained within a single phage particle. If a cell is infected by two genetically different phages, recombination does not occur within segments, but segments may reassociate in forming new particles. This reassociation has the effect of recombining parental genes on different segments. A possible functional explanation of the segmented genome structure (which is not uncommon among RNA viruses⁴) is that it serves to retard Muller's ratchet.

Chao maintained 20 lines of the phage for 40 growth cycles, each cycle involving reduction of the population to a single particle, followed by growth to about 8×10^9 particles. After 40 cycles, the fitness (measured as the rate of multiplication in a paired growth experiment) of these clones was compared to that of a genetically marked parental clone. The transferred clones varied widely in fitness, but the mean value was significantly reduced to 78 per cent of that of the parental clone. Chao interprets this decline as arising from Muller's ratchet.

To evaluate this, we must understand how and when the ratchet operates. Suppose a population consists of N individuals, of which n_0 have no deleterious mutations. If n_0 is a small number, then even though individuals with no mutations are the fittest in the population, it may be that, by chance, all n_0 will die without leaving progeny. If there is recombination, the optimal class can be regenerated, because an individual with no mutations can be produced as a recombinant between two individuals carrying deleterious mutations at different sites. But in the absence of recombination, once the optimal class has been lost it cannot be regenerated (except by the unlikely occurrence of back mutation or compensating mutation): the ratchet has clicked round one notch. This process will occur only in a finite population, but it does not mean that the population must be small. Haigh⁵

showed that, assuming multiplicative fitnesses, $n_0 = N \exp(-U/s)$, where U is the rate of deleterious mutation per genome, and s the selective disadvantage caused by a single mutation. So if the harmful effect of a single mutation is small, the ratchet will operate even in a very large population. However, it will operate more rapidly if the population is small, or, as in Chao's experiment, is periodically reduced to a single individual.

Processes other than Muller's ratchet may have been involved in the deterioration of the transferred lines. Genetic drift (that is, random sampling in a finite population) has two quite distinct effects on the fitness of a finite population. First, it may by chance eliminate the n_0 fittest individuals, which is the process of Muller's ratchet. Second, it disguises fitness differences, so that the effective selective disadvantage, s , caused by a deleterious mutation is smaller than it would be in the absence of drift. If $s < U/\ln N$ in Haigh's model, then the optimal genotype would

not even exist at equilibrium in a population with recombination; that is, n_0 would be less than 1. This is the distinct phenomenon of the 'error threshold' in a finite population⁶. Chao's experimental protocol ensures that any mutant which is capable of forming a detectable plaque at all has the same fitness as the original strain, which means that mutations with small deleterious effects will not be selected against. The process of repeatedly passing each lineage through a bottleneck of any one virus capable of forming a plaque means that the lineage simultaneously faces the twin problems of Muller's ratchet and the error threshold, the latter induced by such extreme genetic drift.

The high mutation rate of RNA viruses, the periodic reduction to a single individual and the absence of reassortment of segments between genetically different viruses would all favour the operation of the ratchet. It is therefore plausible that the decline in fitness of the transferred lines was caused by the accumulation of deleterious mutations. It is harder to decide whether, in nature, the segmental genome structure serves to prevent the ratchet operating. If there is no recombination within segments, the ratchet will

A fresh start for Mount St Helens



EXACTLY 5 months after the cataclysmic eruption of 18 May 1980 removed the top 400 metres of Mount St Helens and the whole of its north flank, leaving a crater 1.5 km across, renewed magmatic activity started the slow reconstruction of the volcano. The lava dome glowing in the centre of the picture first appeared on 18 October of that year (two earlier nascent domes were blown away by explosive eruptions). Then it was around 10 metres high and 25 metres across. At the end of its last active phase in October 1986 (when the photograph above was taken) the dome stood 267 metres above its vent. From photographic monitoring, J.H. Fink *et al.*, on page 435 of this issue, show that the dome grew by a gradual swelling from within (endogenous growth), punctuated by fracturing that allowed lava to flow over the stretched carapace. Although the latter, extrusive mechanism was initially the dominant mode of growth that has since weakened while endogenous growth has continued undiminished. The authors relate this change to the thickening of the dome's crust, which makes fracture less likely. (Photograph by Lyn Topinka, USGS.)

R.P.

still operate in each segment separately, but with a reduction in the relevant mutation rate, U , by a factor of three in the case of $\phi 6$. This may be important: it permits an increase in the total genome size for a given per base mutation rate⁴. It is known that sex will restore the health of a lineage

1. Muller, H.J. *Mut. Res.* **1**, 2–9 (1964).
2. Felsenstein, J. *Genetics* **78**, 737–756 (1974).
3. Chao, L. *Nature* **348**, 454–455 (1990).
4. Pressing, J. & Reaney, D.C. *J. molec. Evol.* **20**, 135–146 (1984).
5. Haigh, J. *Theor. Populat. Biol.* **14**, 251–267 (1978).
6. Nowak, M. & Schuster, P. *J. theor. Biol.* **137**, 375–395 (1989).
7. Bell, G. *Sex and Death in Protozoa. The History of an Obsession* (Cambridge Univ. Press, 1988).

METABOLISM

How to fuel a hummingbird

Jared M. Diamond

FLYING hummingbirds face the problem of fuelling the highest known rates of aerobic metabolism among vertebrates. Migratory hummingbirds have the additional problem of building up large fat deposits (10 per cent of their body mass for each day of premigratory fattening) to meet the energy demands for their long migratory flights. How do they balance the contradictory requirements of fuelling their foraging while conserving fuel for migration? A study by Suarez *et al.*, published in *Proceedings of the National Academy of Sciences* (**87**, 9207–9210; 1990) yields the answer: the birds rely on different energy stores for the two activities.

Hummingbirds' high metabolic rates stem from the fact that they (together with bumblebee bats) are the smallest endothermic vertebrates, and that their preferred foraging mode of hovering is energetically expensive. Suarez *et al.* chose as their study species the Rufous hummingbird, of western North America, which migrates 3,000 km between its summer and winter grounds. To identify the metabolic fuel being used at any instant, Suarez *et al.* measured the respiratory quotient (RQ, the ratio of CO₂ produced to O₂ consumed), which is lower for fat oxidation (0.70) than for carbohydrate oxidation (1.00). The authors induced hummingbirds to hover at an artificial nectar dispenser functioning as a flow-through gas mask, and thereby determined the RQ during a single bout of foraging.

Fasted resting birds yielded an RQ of 0.72, showing that fat is then the main fuel. The RQ rose to 0.81 during the first bout of hover-feeding after a fast, then to 1.0 during subsequent bouts. These values mean there is a shift in energy source from fat to carbohydrate during foraging.

Why? The reason is that hummingbirds must build up fat stores for migration by synthesizing fatty acids from the sugar

of protozoa that is senescing after a period of enforced chastity⁷. It would be interesting to know whether, in an experiment similar to that described by Chao, the transferred lines would still decline in fitness if there was an occasional opportunity for segment reassortment between different lines, perhaps at the end of each growth cycle. □

John Maynard Smith is in the Department of Biology, University of Sussex, Brighton BN1 9QG, UK. Sean Nee is in the AFRC Unit of Ecology and Behaviour, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

ingested in nectar. Burning the sugar itself during foraging not only spares fat stores but also avoids the inefficiency of synthesizing fatty acid from glucose and then burning the fat. Migrating hummingbirds, however, have no choice except to burn fat, because of its higher calorie yield per

top all other animals by large margins in their rates of intestinal glucose absorption, hepatic fatty acid biosynthetic capacity (over ten times that of mammals), and muscle levels of hexokinase and carnitine palmitoyltransferase. Comparison of the capacity of these latter two enzymes for glucose and fatty acid oxidation, respectively, with actual fuel consumption rates during flight shows a capacity to correspond closely to consumption — that is, there is little excess capacity or safety margin.

This comparison is all the more interesting in light of current discussions of the evolutionary design of metabolic machinery. C.R. Taylor and E. Weibel (*Respiration Physiol.* **44**, 1–10; 1981) coined the term 'symmorphosis' to refer to the concept that linked steps in physiological systems should have capacities matched to each other and to prevailing natural loads — if design efficiency were the overriding criterion. Such a framework begs the question of how much safety margin, if any, is designed into physiological systems. Typically, there seems to be a 100 per cent safety margin for nutrient transporters in the mammalian



B & C Calhoun/Bruce Coleman

The hummingbird — hovering close to the metabolic limit.

gram than carbohydrate and because of the need to save weight in flight.

Hummingbird liver and flight muscle contain enough glycogen to support metabolism for only five minutes of flight. By keeping bouts of foraging well under this time limit, the birds can fuel their foraging by glycogen reserves that they replenish from sugar, and can spare their fat stores. Their use of fat after a long fast may be necessitated by exhaustion of glycogen.

Supporting these record-high metabolic rates requires hummingbirds to be jam-packed with metabolic machinery. They

intestine (R. Ferraris and J. Diamond *A. Rev. Physiol.* **51**, 125–141; 1989), but very little if any for the hummingbird enzymes studied by Suarez *et al.* Perhaps hummingbirds have reached a design limit imposed by the difficulty of packing any more enzymes into their flight muscles, where conditions must already have virtually reached standing-room-only at the molecular level. □

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90024, USA.

One hand clapping

Paul Travers

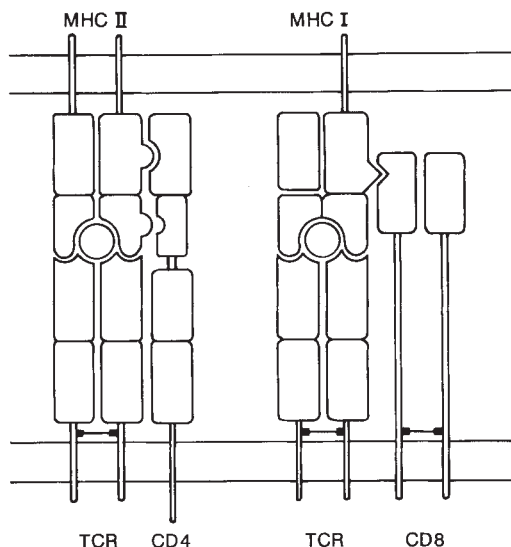
ALTHOUGH the CD4 molecule is notorious as the cellular receptor for the human immunodeficiency virus (HIV)¹, its function in the normal physiology of the cellular immune system is as a receptor for major histocompatibility complex (MHC) class II antigens², an accessory to the T-cell receptor³ and a receptor for p56^{lck}, a cellular tyrosine kinase⁴. The determination of the structure of a soluble fragment of the CD4 antigen, reported in this issue by Wang *et al.*⁵ on page 411 and Ryu *et al.*⁶ on page 419, clarifies to some extent the interaction of CD4 with HIV and with class II antigens. As a bonus, the molecule provides an example of a new class of immunoglobulin-like domains and a novel mode of interaction between immunoglobulin domains, one that is surely to be found in other superfamily members⁷.

There are four external immunoglobulin-like domains (V1 to V4) in the CD4 molecule and there has been some debate as to whether these domains interact to form a compact globular structure or are strung out in a linear molecule, looking, for example, like an isolated immunoglobulin heavy chain. Some support has been given to the linear model by crystallographic studies that have suggested that the complete molecule forms a long, flexible rod^{8,9} about 125 Å in length. Unfortunately, the data obtained from crystals of the four-domain protein are not sufficient to enable a molecular structure to be derived.

The structures of engineered fragments containing the two N-terminal domains that have now been produced are consistent with an overall linear molecule but suggest the whole protein may resemble a flail — two rigid rods with a flexible connection — rather than a string of beads. The V1V2 fragment forms a rigid rod of about 60–65 Å in length, comprising roughly half of the complete molecule. The rigidity of the molecule is due mainly to lack of a spacer or connecting peptide linking the V1 and V2 domains. Instead, a long β strand forms the last strand of the V1 domain and continues as the first strand of the V2 domain, creating a central 'spine' from which the two domains are suspended. Such a rigid connection between domains may well be a feature of other immunoglobulin superfamily proteins, particularly those that, like CD4, are monomeric. The platelet-derived growth factor (PDGF) receptor, neural cell adhesion molecule (NCAM) and vascular cell adhesion molecule (VCAM) are all potential candidates for this type of rigid structure.

The sequence homologies between immunoglobulin and the V1 and V2

domains have been used to predict the structures of those domains. Models of the V1 domain, based on an alignment with the REI Vκ domain, were used to interpret the results of mutagenesis experiments that mapped the binding site for the HIV envelope glycoprotein, gp120 (ref. 10). The site was localized to a region in CD4 corresponding to an immunoglobulin CDR2 loop. The structures now



The CD4 molecule can span the T-cell receptor (TCR)–MHC complex: for the CD8 molecule to do so requires an extended conformation of part of its polypeptide chain.

obtained show that the conformation of this loop, as well as others in the V1 domain, differs from that in immunoglobulin variable (V) domains. The CDR2-equivalent loop in CD4 is longer than that in immunoglobulins, and an extended core of hydrophobic residues pushes it away from the rest of the domain to create a prominent ridge. Most of the residues critical in gp120 binding lie in this ridge, which is thought to interact with a complementary groove in the gp120 surface. The ridge is the obvious target for competitive inhibitors of binding, and the idea that binding may involve a groove in gp120 rather than an extended surface surely makes the design of such inhibitors more credible. Further, the identification of a second site on the V1 domain, located in a cluster of negatively charged residues, that is implicated in viral fusion with host cell membranes provides an additional target for inhibitors of the viral life cycle.

In contrast to the predictions for V1, the prediction for V2 (ref. 7) correctly identified the main differences between the V2 domain and immunoglobulin V and constant (C) domains. The V2 domain is truncated relative to both immunoglobulin V and C domains and has a novel strand

topology (see Fig. 3 of Ryu *et al.*⁶). The most unusual feature of this domain is that the disulphide bridge links adjacent strands rather than crossing between two β sheets. Several immunoglobulin superfamily proteins have domains that are thought to have non-typical disulphide bridges (CD8, ref. 11) or lack the disulphide bridge altogether (CD4 domain 3, ref. 7; PapD, ref. 12). The structures of the V2 domain and the bacterial chaperone, PapD, show how the general features of the immunoglobulin fold can be maintained without the disulphide bridge. As far as modellers are concerned,

the immunoglobulin superfamily building kit now contains three types of bricks and a new set of rules on how to put them together. But the failure to predict the loop differences in the V1 domain and the rigid connection between the two domains is a warning that a simple extrapolation from known structures is unlikely to uncover the subtle variety of evolution. The success of one group in predicting the salient features of the V2 domain reflects, I think, the success of a pragmatic methodology over more theoretical approaches.

The 'real' functions of the CD4 molecule involve the binding of MHC class II antigens, T-cell receptors and p56^{lck}. Of these, only the binding of class II proteins is likely to be illuminated by the structures that have now been solved. Residues in both the V1 and V2 domains affect class II binding and these are distributed on either side of a concavity or depression formed by the packing of the V1 and V2 domains. It has been suggested that the binding site for pili subunits on the PapD molecule also lies in such a depression between two longitudinally interacting immunoglobulin domains. Thus the ligand is cupped in the palm of one hand rather than cradled between two; the creation of a binding site by lateral interactions between domains, as in immunoglobulins themselves, may prove to be an exception rather than the rule.

Unfortunately, the binding site for CD4 on the MHC class II molecule is not known, so the degree of complementarity of the CD4 and class II sites cannot be assessed. The binding site on MHC class I for the analogous T-cell accessory molecule, CD8, is known to lie in the α3 domain that lies near the membrane. The length of an MHC molecule, about 70 Å, is roughly the same as that of the CD4 V1V2 fragment and it is possible that CD4 is in contact with both membrane-proximal and membrane-distal domains of class II antigens. The T-cell receptor is probably about 60–70 Å long (or is it tall?), so the complete CD4 molecule, at 125 Å or so,

could span from the T-cell membrane to the membrane-proximal domains of an MHC class II molecule interacting with a T-cell receptor¹³, as shown schematically in the figure. Such a model implies an even more unusual structure for the CD8 molecule. The CD8 monomer has a single V-like domain linked to the membrane by about 40 amino acids, which would have to be in a relatively extended conformation to allow the molecule to span the 90 Å or so required; the sensitivity of CD8 to cleavage by trypsin¹⁴ suggests that this might be so.

Whatever impact solving the structure

of these two domains has on the search for inhibitors of HIV, it is clear that the structure of the CD4 V1V2 fragment provides new ways of looking at immunoglobulin superfamily proteins. With new domain structures, new interdomain interactions and a new model for the construction of ligand-binding sites, I foresee models aplenty for this ever-growing family of proteins — at least, until the next structure comes along. □

Paul Travers is in the Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HE, UK.

1. Dalgleish, A. *et al.* *Nature* **312**, 763–766 (1984).
2. Doyle, C. & Strominger, J. L. *Nature* **330**, 256 (1987).
3. Chuck, R. S., Cantor, C. R. & Tse, D. B. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5021–5025 (1990).
4. Rudd, C. E. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 5190–5194 (1988).
5. Wang, J. *et al.* *Nature* **348**, 411–418 (1990).
6. Ryu, S. E. *et al.* *Nature* **348**, 419–426 (1990).
7. Williams, A. F. *et al.* *Cold Spring Harb. Symp. quant. Biol.* **54**, 637–647 (1989).

8. Davis, S. J. *et al.* *J. molec. Biol.* **213**, 7–10 (1990).
9. Kwong, P. D. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **87**, 6423–6427 (1990).
10. Arthos, J. *et al.* *Cell* **57**, 469–481 (1989).
11. Kirsbaum, L. *et al.* *J. Immun.* **142**, 3931–3936 (1989).
12. Holmgren, A. & Branden, C. I. *Nature* **342**, 248–251 (1989).
13. Davis, M. M. & Bjorkman, P. J. *Nature* **334**, 395–401 (1988).
14. Ledbetter, J. A. *et al.* *J. exp. Med.* **153**, 1503 (1981).

MANTLE GEOCHEMISTRY

Reducing mantle redox options

Simon L. Harley and Nicholas W. A. Odling

IT is widely accepted that volatile species in the carbon–oxygen–hydrogen system dominate the fluid-present processes in the upper mantle¹. The redox state of the mantle controls the fluid speciation and hence the types of magma produced at depth: both the percentage melting at a given pressure and temperature and the composition of the melts are strongly influenced by whether, for example, oxidized (H₂O–CO₂) or reduced (CH₄–H₂O) fluids have induced melting^{2–4}. Clearly, in the present Earth, there are large-scale (tectonic) and small-scale processes which could influence the oxygen fugacity of the mantle and hence the melting relations. Any demonstration of a correlation between magmatism, oxygen fugacity and tectonic processes, such as that presented on page 437 of this issue⁵ by Ballhaus *et al.*, is therefore of fundamental significance for the modelling of plate processes, interactions between the asthenosphere and lithosphere, and the evolution of the crust–mantle system.

Until only recently, arguments over the existence of f_{O_2} variations between mantle samples and melt products were hampered by a lack of agreement between the calibrations of the indirect methods, such as f_{O_2} -sensitive reactions^{5–8}, at our disposal for measuring f_{O_2} . This disagreement has stemmed partly from an absence of comprehensive experimental data in the pressure–temperature–composition– f_{O_2} ranges appropriate to the upper mantle. It is an oxygen barometer based on such experimental study that Ballhaus *et al.* now present. This study allows critical comparisons to be made between the

earlier, largely indirect methods; it shows encouraging agreement with the most recent of these calibrations⁹. As a consequence, we now have significantly greater confidence that the consistent variations in f_{O_2} monitored by upper-mantle mineral assemblages can be reliably quantified and that the implied range of three to four log units in mantle f_{O_2} is indeed real and not simply an artefact of the precision of the calibrations.

Thus the f_{O_2} populations defined by Ballhaus *et al.*, and also noted by Wood *et al.*⁹, indicate that there are significant large-scale variations in the upper-mantle redox state, and that distinct f_{O_2} regimes may be associated with specific tectonic environments. The clearest example of this is the highly oxidized condition of island arc basalts (IAB), generated at subduction zones, which may be correlated with processes that cause enrichment in the light rare-earth elements (LREE) and large-ion lithophile elements (LILE), currently thought to be important in the genesis of subduction-related magmatism. This oxidized condition contrasts with the relatively reduced states implied for primitive mid-ocean ridge basalts (MORB) and the possibly conjugate, relatively undepleted (primitive) abyssal peridotites.

There remains some disagreement as to the redox state of the processes responsible for hotspot (or ocean island) basaltic magmatism. In their analysis, Ballhaus *et al.* show that xenoliths in ocean island basalts (OIB) are relatively oxidized and hence argue that the lithospheric mantle affected by hotspot plumes is also oxidized. Wood *et al.*⁹, in contrast, have suggested

that some reduced abyssal peridotites spatially associated with plume-related basalts could imply that those mantle plumes are more reduced than 'normal asthenospheric mantle'. Those samples in which a 'plume' signature can be identified are of vital importance as they present a window through which we may be able to examine the processes of lithospheric/asthenospheric interaction commonly believed to be associated with upper-mantle geochemical enrichments. It must be recognized that such samples may well be composites including fractions derived from both the lithosphere and asthenosphere, and that decompression and possible melt extraction could lead to further modifications of the f_{O_2} signatures. Both Ballhaus *et al.* and Wood *et al.* emphasize the extreme sensitivity of reduced mineral assemblages to redox perturbation by interaction with oxidized material. As a consequence, there is considerable uncertainty inherent in the use of xenolith samples for inferring the redox state of deep-mantle plume material responsible for the geochemical enrichments.

Clearly, the present confidence in our ability to determine the f_{O_2} established by the calibrations of Ballhaus *et al.* extends only as far as the limits of the spinel stability field in the mantle. The extrapolations of low-pressure electrochemical data necessary to extend our understanding of the redox structure beneath the spinel stability field may result in calibrational uncertainties which prohibit recognition of any real redox variations at greater depths in the mantle. To reach an adequate understanding of the redox condition of the deeper levels of the upper mantle, and hence address the questions of the redox states of plumes and thickened lithospheres, it is essential that we formulate and calibrate f_{O_2} -sensitive reactions involving Fe²⁺/Fe³⁺ relations in appropriate high-pressure garnet-bearing assemblages⁷ analogous to the low-pressure spinel-bearing assemblage calibrated by Ballhaus *et al.*⁵. The prime component of such a calibration must be an experimental study covering a range in f_{O_2} conditions and within the pressure–temperature–composition field appropriate for the upper mantle. □

Simon L. Harley and Nicholas W. A. Odling are in the Grant Institute of Geology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JW, UK.

1. Eggler, D.H. in *Mantle Metasomatism* (eds Menzies, M.A. & Hawkesworth, C.J.) 21–41 (Academic, London, 1987).
2. Holloway, J.R. *Adv. Phys.* **1**, 273–293 (1981).
3. Falloon, T.J. & Green, D.H. *Earth Planet. Sci. Lett.* **94**, 364–370 (1990).
4. Taylor, W.R. & Green, D.H. *Nature* **332**, 349–352 (1987).
5. Ballhaus, C., Berry, R.F. & Green, D.H. *Nature* **348**, 437–440 (1990).
6. Arculus, R.J. & Delano, J.W. *Geochim. cosmochim. Acta* **45**, 889–913 (1981).
7. Mattioli, G. & Wood, B. *Nature* **322**, 626–628 (1987).
8. O'Neill, H.St.C. & Wall, V.J. *J. Petrology* **28**, 1069 (1987).
9. Wood, B.J., Bryndzia, L.T. & Johnson, K.E. *Science* **248**, 337–345 (1990).

Eating people is wrong

Paul G. Bahn

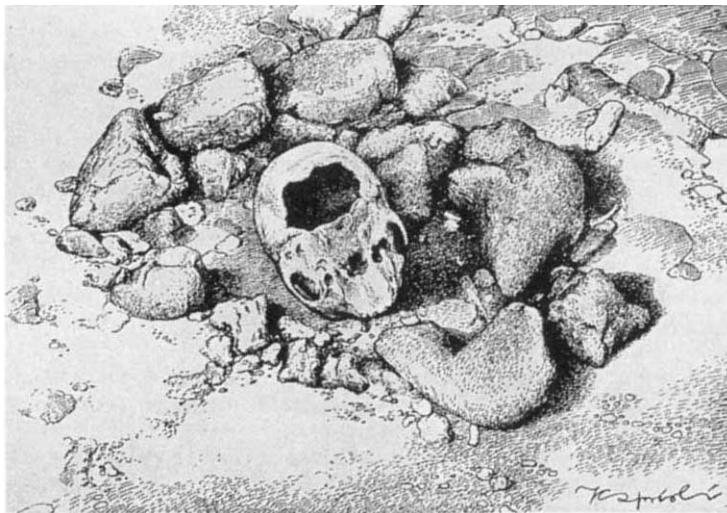
THE notion that early humans were cannibalistic may at last have been laid to rest following a re-evaluation of the famous Neanderthal skull from Monte Circeo, Italy. Once thought to be firm evidence of ritual cannibalism, new studies show that the skull was in fact gnawed by hyaenas^{1,2}. The work comes at a time when much evidence for prehistoric cannibalism is being revealed as the result of wishful thinking imposed on bone assemblages of usually rather more mundane provenance³⁻⁷.

The skull (of a male, about 45 years old) was discovered in 1939 in a cave in the Monte Circeo, 100 km south of Rome. Hours after its initial discovery, the prehistorian Alberto Blanc took it to Rome without having taken any photographs of the skull *in situ*. He later produced a drawing (see figure) showing the skull, base-upwards, in a 'crown of stones', a feature subsequently revealed as spurious. The picture disguises the fact that the discoverer had picked the skull up and replaced it on the floor before Blanc ever saw it, and that Blanc was unsure about the original position of the skull.

A blow to the right temple was thought to have been the cause of death, and the enlarged hole in the base was taken to be evidence for the extraction of the brain for consumption. This diagnosis, together with the cranium's position in a ring of stones, strongly suggested ritual cannibalism. Innumerable popular works have since accepted this thesis without question, and there has even been speculation, because of its position, that the skull was used as a drinking vessel. But the pattern of colouring and concretions on the cranium indicate that it originally lay on its left side on the floor².

Moreover, the skull was just one of many bones scattered all over the cave,

representing a wide variety of species, with a high ratio of carnivores to herbivores and low degree of fragmentation. Juveniles are rare, as are vertebrae and ribs, whereas limb bones are abundant. All of these factors have since been identified as characteristic of hyaena den



Misleading records: the picture of the Monte Circeo skull, drawn from memory.

assemblages, as are diagnostic impressions left by hyaena teeth on gnawed bone^{8,9}. Blanc himself identified carnivore toothmarks on some of the bones¹⁰, and the presence of hyaena bones and coprolites (fossil faeces) led him to conclude that tenancy of the cave alternated between hyaenas and humans. The skull itself shows no sign of being modified by human agency^{1,2}: the temporal fractures are consistent with hyaena toothmarks, and the enlarged hole at the base has been

gnawed round the edges. The only other human bone found at the site, a jawbone, is almost certainly from the same individual² and has also been gnawed by hyaenas. Dating of bones from the cave floor shows that the hyaenas were present about 52,000 years ago.

Cannibalism has been a favourite theme in archaeology for a long time, and is a preoccupation that recurs up to the present day. For example, the bones of several dozen human individuals from a 6,000-

year-old neolithic cave site at Fontbrégoua in France were recovered from disposal pits that also contained animal bones. They bore cut marks and breakage patterns identical to those on animal remains, and were interpreted as the left-overs from cannibalistic defleshing of fresh bones¹¹. But mortuary rituals among some Australian Aborigines can produce precisely this pattern¹ after bodies have been left exposed for a while to decompose, and have then been defleshed for secondary disposal. So for the time being, at least, the case for cannibalism at Fontbrégoua, and other sites

where it has been invoked (such as the recently reassessed 'cannibal feast' of Neanderthal bones at Krapina, Yugoslavia¹²) must now be judged 'not proven'. Interestingly, the only concrete proof of cannibalism — the presence of human remains in human coprolites — has never been found anywhere. □

Paul G. Bahn is a freelance writer on archaeology based at 428 Anlaby Road, Hull HU3 6QP, UK.

ASTRONOMY

The Galaxy's turbulent youth

M. G. Edmunds

VIOLENT periods in history always make for more interesting reading than the quiet years. For this reason, recent investigations¹⁻⁴ of the early history of our Galaxy suggesting that its formative years may have been longer and more chaotic than previously realized, make a welcome change.

A conventional view of the evolution of the Galaxy supposes the fairly rapid formation of a non-rotating (or at least, only slightly rotating) roughly spherical distribution of stars — the halo — followed by collapse of the bulk of the primordial gas to a thin disk in which stars continue to form to this day. It was long believed that the formation of the halo was sufficiently fast, perhaps less than a

billion years out of the Galaxy's total age of 12–15 billion, that the disk essentially evolved as a separate system. During the 1980s deep star-count surveys began to indicate that there might be an intermediate thick disk, presumably formed both in space and time between the halo and the thin disk. And now it may be necessary to abandon even this simple sequential picture, halo → thick disk → thin disk.

One line of evidence comes from a comparison of the ages of globular clusters. Because these star clusters seem to have the same kinematics and range of chemical abundances as the halo stars, it is reasonable to assume that they are a part of the halo. Despite prodigious effort, it has proved very difficult to obtain

1. Piperno, M. & Giacobbia, G. *Quaternaria Nova* (in the press).
2. Giacobbia, G. & Piperno, M. In *The Circeo 1 Neanderthal Skull. Studies and documentation*, 316–322 (Museo Nazionale Preistorico Etnografico L. Pigorini, Rome, 1990).
3. Arens, W. *The Man-Eating Myth: Anthropology and Anthropophagy* (Oxford University Press, 1979).
4. Pickering, M. P. *Aust. Archaeol.* **28**, 35–39 (1989).
5. Trinkaus, E. *J. hum. Evol.* **14**, 203–216 (1985).
6. Russell, M.D. *Am. J. phys. Anthropol.* **72**, 373 (1987).
7. Russell, M.D. *Am. J. phys. Anthropol.* **72**, 381–397 (1987).
8. Sutcliffe, A.J. *Nature* **227**, 1110–1113 (1970).
9. Brain, C.K. *The Hunters or the Hunted? An Introduction to African Cave Taphonomy* (Chicago Univ. Press, 1981).
10. Blanc, A.C. *Materie Prime d'Italia e dell'Impero* **7–8**, 3–23 (1939).
11. Villa, P. *et al. Science* **233**, 431–437 (1986).

accurate absolute ages for globular clusters to test this assumption — which is also unfortunate for cosmology, as these objects provide the oldest identifiable structures in the Galaxy and would provide a good lower bound on the age of the Universe.

The best attempts can be made by comparing observations of the brightness and colours of their stars with theoretical models of stellar evolution, but the estimates are uncertain because of the influence of the chemical composition of the stars on their evolution. The chemical content affects lifetimes both by regulating the transfer of radiation through the star and in determining how much nuclear fuel is available. Nevertheless, by comparing globular clusters which have the same chemical composition¹, VandenBerg *et al.* have obtained *relative* ages of some clusters to an accuracy of half a billion years.

The interesting result is that for the chemically poorest (and, hence, probably the oldest) clusters, there is very little dispersion in age, whereas the more chemically rich clusters may have formed over an extended period in excess of two billion years. This implies a rapid, coherent start to the halo formation, but that the formation of the clusters then became more sporadic while the chemical abundances built up as a result of supernova explosions of the massive stars. The star clusters may have formed here and there in regions of different chemical composition, although locally the regions of cluster formation must have been chemically homogeneous because (with few exceptions) the stars in each cluster share a common chemical abundance.

Meanwhile the disk was forming. The most obvious difference between the disk and the halo (apart from chemical differences) is in their rotational properties. The disk rotates rapidly, whereas the average rotation of the halo is small, with some stars even on contra-rotating orbits. Presumably the disk's fast rotation is a result of conservation of angular momentum during the collapse associated with its formation. Stars with chemical abundances below about a sixth of that found in our local thin-disk star, the Sun, have generally been found to belong to the slowly-rotating halo. At least two surveys^{2,3} now find stars which have considerably lower metal abundances (1/40–1/10 of solar) but whose rapid rotation about the Galactic Centre implies that they are true disk stars, probably of the thick component. Thus it seems likely that star formation in the disk started not long after the start of halo formation, and both proceeded together for some considerable time.

Two other, rather preliminary results are rather curious. In the work by Morrison *et al.*², there is an indication that the

fall off in density of stars with distance from the Galactic Centre may be steeper in the thick disk than in the thin component. This is the reverse of what might be expected if the thin disk had formed by contraction inside the thick disk. Second, although it has been known for some time that many red giant stars near the centre of the Galaxy have a high chemical abundance — a metal-rich component of the inner region of the halo — it has been assumed that the halo out near the Sun is chemically poor. Infrared photometry of M-type giant stars⁴ now suggests that there are some M giants high in the halo above us which are just as chemically rich as stars in the nuclear region. These cool M giant stars have molecule-filled atmospheres which are rather tricky to analyse accurately, so it may be premature to accept the existence of an extended chemically rich halo component. But if it is real, this could be interpreted as further evidence indicating an extended period of star formation in the halo. It is hard to see how the old-disk stars could have low chemical abundance if a preceding halo had built up high abundances, unless there were very large chemical inhomogeneities.

Whether halo star formation continued until quite recent epochs is difficult to say, but it is an important question. If it did, there would be considerable implications for observational searches for evolutionary changes in other galaxies at relatively modest redshifts. A good test would be the discovery of stars showing halo-type kinematics but which are too massive, and hence too short-lived to have survived since the early epoch of halo formation. Metal-rich F or early G-type dwarfs would be suitable. Possible halo candidates with considerably higher masses are known, but they are generally accounted for in other, slightly *ad hoc* ways.

Faced with these latest archaeological artefacts, it may be prudent to abandon fixed ideas of a fast Galaxy formation. As realized by the authors of these new studies, the real formation may have been a much more drawn out and disorganized process⁵, with isolated regions evolving on their own for some time before merging into the fairly well-defined structure we see. How long these dark ages actually lasted is a crucial question to answer, and fixing the absolute times at which the globular clusters and halo stars formed would be a significant advance. □

M. G. Edmunds is in the Department of Physics, University of Wales College of Cardiff, PO Box 913, Cardiff CF1 3TH, UK.

1. VandenBerg, D. A., Bolte, M. & Stetson, P. B. *Astr. J.* **100**, 445–467 (1990).
2. Morrison, H. L., Flynn, C. & Freeman, K. C. *Astr. J.* **100**, 1191–1222 (1990).
3. Norris, J., Bessell, M. S. & Pickles, A. J. *Astrophys. J. Suppl.* **58**, 463–492 (1990).
4. Feast, M. W., Whitelock, P. A. & Carter, B. S. *Mon. Not. R. astr. Soc.* **247**, 227–236 (1990).
5. Searle, L. & Zinn, R. *Astrophys. J.* **225**, 357–379 (1978).

Firefighting foam

THE greenhouse effect, that increased retention of the Sun's radiation by carbon dioxide accumulating in our atmosphere, is now worrying even politicians. One possible answer, which has been dignified by editorial discussion in these pages, is to reflect sunlight away from the Earth by terrestrial or space-borne mirrors. Even Daedalus rejects this notion as sheer technomania; but the idea behind it is sound. Extra cloud in the atmosphere would have the same effect. Does the vapour trail of an aircraft reflect away enough sunlight to cancel the effect of the extra carbon dioxide it puts into the air? Sadly, the answer is no; but a related idea looks much more hopeful.

A ship can leave a turbulent wake hundreds of metres wide, and thousands of kilometres long. If the white foam produced by its turbulence could be stabilized in some way, it could reflect many gigawatts of solar radiation away from the Earth for as long as it lasted. Daedalus recalls the early, non-biodegradable domestic detergents, and the terrible pollution they caused in rivers and sewage works with their tenacious, almost permanent foam. DREADCO's chemists are now resurrecting some of these long-banned detergents, and trying to make them more surface active, polluting and permanent still. A sufficiently surface-active detergent would not dissolve in bulk seawater at all, but would spread over the surface as a sort of slick. Spread from a ship, it would convert the transient wake into a wide swathe of stable white foam. Such a foam would reflect maybe 80 per cent of the sunlight hitting it, instead of the 5 per cent or so of normal seawater. A good detergent can foam in such low concentrations that a few tons of it, leaked steadily into the sea by a transatlantic vessel, might leave a stable foamy track the whole way across.

If sufficiently biologically inert and polluting, such a foam could last for weeks or months. Even when it finally collapsed, the underlying detergent slick would still remain, ready to be whipped into foam again by storms, or further passing ships. Daedalus estimates that the world's shipping could easily keep 5–10 per cent of the ocean surface covered with reflecting foam, quite enough to cancel out the greenhouse warming of our industrial and forest-burning carbon dioxide emissions.

Thus the greenhouse can be staved off without giving in to the Greens. All that is required is a simple modification to the world's shipping. Carbon taxes, limits to growth, conservation, harmony with nature, and pettifogging searches for efficiency and economy, can all be quietly forgotten. The glorious expansionist machismo of industrial growth and profligacy will be able to roll grandly on till Kingdom come.

David Jones

Land use in the Amazon

SIR—There have been several recent analyses^{1,2} of the actual and potential market values of intact forest in the Peruvian Amazon, in which a future benefit analysis approach to forest valuation is emphasized. But socio-economic reality in the region contradicts assumptions implicit in future benefit and net-present-value analyses. Rural inhabitants of northeast Peru use their land for a variety of subsistence and market-oriented uses: some are based on slash-and-burn conversion of forested land to agriculture and others on extraction of forest products. Slash-and-burn agriculture has exceeded extraction of forest products in economic importance since the 1950s (ref. 3). The hundreds of thousands of shifting cultivators inhabiting the Peruvian Amazon are not misguided recent immigrants, yet their decisions about land use are neither governed nor explained by calculations of perpetually accruing net revenues.

We suggest three reasons for this situation. First, predictions of future benefits rely on estimates of future commodity prices. For such forecasts to be reasonably accurate, prices must be relatively stable or at least relatively predictable; prices for perishable goods with limited markets (forest fruits in Iquitos, Peru, for example) are neither⁴. In the Peruvian Amazon, markets for agricultural staples are more dependable than markets for non-timber forest products, presenting a lower risk for producers. Prices for rice are set by the government, and credit is available for the production of rice, cassava and plantains, providing further incentive for their cultivation (ref. 5 and M. Chibnick, personal communication).

Second, most individuals and rural communities in the Peruvian Amazon lack secure land and resource tenure. As a result, there is little incentive to pursue options for land use with immediate returns lower than those available from slash-and-burn agriculture, regardless of potentially realizable future benefits.

Third, decisions about use of land in the region are based on both subsistence and market-oriented considerations. Net-present-value calculations often neglect the value of subsistence activities because they are difficult to measure.

Economic analyses of alternative uses of land in the Peruvian Amazon must reflect the context of the choices faced by rural populations. Widespread conversion of intact forest to agricultural use is

neither ecologically nor, in the long term, socially desirable. Nonetheless, rural populations can be expected to continue converting forested land to slash-and-burn agriculture unless alternative land uses become more attractive in practice as well as in theory.

MIGUEL PINEDO-VASQUEZ
DANIEL ZARIN
PETER JIPP

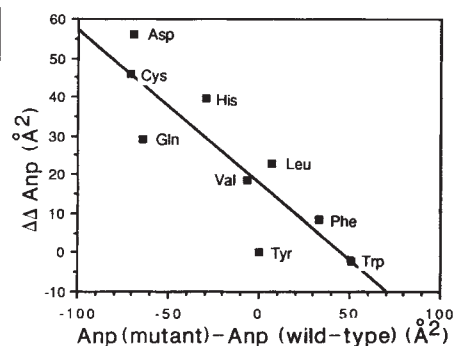
School of Forestry and
Environmental Studies,
Yale University, New Haven,
Connecticut 06511, USA

Surface areas of unfolded proteins

SIR—Pakula and R.T.S. have recently investigated¹ the effects of amino-acid substitutions at position 26 in the protein λ Cro. This residue is unusual in that it is more exposed to solvent in the folded than the unfolded states^{2,3}. Hence it shows a 'reverse' hydrophobic effect as mutations that increase in hydrophobicity decrease in protein stability. The effect of the mutation on ΔG of unfolding correlated with side-chain hydrophobicity, measured by solvent-transfer experiments¹. We have used these results to determine the surface area at position 26 in the unfolded Cro protein.

The solvent-accessible surface areas of residues are usually calculated relative to the residue in an Ala-X-Ala tripeptide^{1,4}. Hence, a side chain with a fractional solvent accessibility of 1.0 has the same solvent-accessible surface area as the side chain in the tripeptide. The mutation studied (Cro, position 26) has an unusually high fractional solvent accessibility of 1.4 in the folded state. The hydrophobicity of a side chain, measured as a free energy, is directly proportional to the non-polar surface area of the side chain⁵, with a constant of proportionality of 24 cal $\text{\AA}^2$. The effect of each mutation on ΔG ($\Delta\Delta G$) can therefore give the change in non-polar surface area ($\Delta\Delta\text{Anp}$) of unfolding between the mutant and wild-type protein, if we assume that the mutations affect ΔG only through the hydrophobic effect.

This assumption is supported by the observed correlation between $\Delta\Delta G$ and hydrophobicity¹. As λ Cro denatures from a dimer to two unfolded monomers¹, each $\Delta\Delta G$ value includes contributions from two mutant side chains. The change in non-polar surface area per mutant side chain obtained from the experiment is therefore $\Delta\Delta G/48$. The Tyr 26 residue is distant from the dimer interface^{2,3} so the folded dimer/folded monomer transition will have no effect on $\Delta\Delta G$. The change in non-polar surface area is equal to ΔAnp



Relationship between the difference in non-polar surface area of unfolding between mutant and wild-type proteins ($\Delta\Delta\text{Anp}$) and the difference in non-polar surface area between the mutant and wild-type residues in Ala-X-Ala tripeptides. The line of best fit is: $\Delta\Delta\text{Anp} = -0.394(\text{Anp}[\text{mutant}] - \text{Anp}[\text{wild-type}]) + 17.7$. Correlation coefficient, 0.87.

between the mutant and wild-type unfolded protein minus ΔAnp between the mutant and wild-type folded protein. The latter is equal to 1.4 multiplied by the difference in non-polar surface area between the mutant and wild-type residues in Ala-X-Ala tripeptides. This assumes that the fractional solvent accessibility of 1.4 for the folded protein is applicable to all the mutations studied. Similarly, ΔAnp between the mutant and wild-type unfolded protein is equal to the fractional accessibility of the unfolded protein multiplied by the difference in non-polar surface area between mutant and wild-type residues in Ala-X-Ala tripeptides.

If we plot $\Delta\Delta\text{Anp}$ (calculated as $\Delta\Delta G/48$) against the difference in non-polar surface area between the mutant and wild-type residues in Ala-X-Ala tripeptides (measured using data in ref. 4), we obtain the difference in fractional accessibilities between the unfolded and folded protein as the gradient. The graph obtained has a slope of -0.39 , showing that the fractional solvent accessibility of position 26 in unfolded λ Cro is $1.4 - 0.39 = 1.01$, a value very close to that expected based on an Ala-X-Ala tripeptide. This suggests that this tripeptide is a reasonable model for the side chains in an unfolded protein, at least for position 26 of λ Cro.

ANDREW J. DOIG
DUDLEY H. WILLIAMS

Cambridge Centre for
Molecular Recognition,
University Chemical Laboratories,
Cambridge CB2 1EW, UK

ROBERT T. SAUER

Department of Biology,
Massachusetts Institute of Technology,
77 Massachusetts Avenue, Cambridge,
Massachusetts 02139, USA

- Peters, C., Gentry, A. & Mendelsohn, R. *Nature* **339**, 655–656 (1989).
- Myers, N. J. *trop. Ecol.* **4**, 209–221 (1988).
- San Román, J. *Perfiles Históricos de la Amazonia Peruana* (Ediciones Paulinas, Lima, Peru, 1975).
- Padoch, C. *Adv. Econ. Bot.* **6**, 214–224 (1988).
- Hiraoka, M. in *Fragile Lands of Latin America: Strategies for Sustainable Development* (ed. Browder, J.O.) 75–101 (Westview, Boulder, 1989).

- Pakula, A.A. & Sauer, R.T. *Nature* **344**, 363–364 (1990).
- Anderson, W.F. *et al. Nature* **290**, 754–758 (1981).
- Ohlendorf, D.H., Anderson, W.F., Fisher, R.G., Takeda, Y. & Matthews, B.W. *Nature* **298**, 718–723 (1982).
- Lee, B.K. & Richards, F.M. *J. molec. Biol.* **55**, 379 (1971).
- Chothia, C. *Nature* **248**, 338–339 (1974).

Population size bottleneck

SIR—Maynard Smith recently estimated¹ the average human population size since the time of the mitochondrial “Eve”. He pointed out that the observed rather small divergence among human mitochondrial

have attained any size and still have accumulated mitochondrial diversity at the same rate.

The risk posed to the survival of the population in the first case is equivalent to the risk of crossing Niagara Falls on a tightrope. In the second case, the risk is more like that posed by balancing for a moment on the sharp end of an ocean liner

Here, k is the Boltzmann constant, h Planck's constant, m the molecular mass, N Avogadro's number, V the molar gas volume, and δ a characteristic transition state distance.

The equivalence of functions 2 and 4 is realized, following Eyring, by substituting the equality $\delta = 1/\nu (kT/2\pi m)^{1/2}$, where ν is the liquid/gas transition-state frequency.

$\Delta S_b^\ddagger$ is then simply evaluated from partition functions 2 = 4, and 3 as $R \ln(V/N\delta^3)$, which can be expressed as

$$R \ln (22,410 \times T / 1.5 \times 273)$$

assuming that $N\delta^3$ is a characteristic volume independent of T with a value of 1.5 cm^3 . This volume is the maximum free molar volume of the ideal liquid in the transition state at its boiling point, so the main result emerging is a specification of the minimum separation (characteristic distance) at which the physical forces between the molecules approximate to zero.

J.J. ROONEY

*School of Chemistry,
Queen's University of Belfast,
Belfast BT9 5AG, Northern Ireland*

1. Maddox, J. *Nature* **345**, 119 (1990).
2. Rooney, J.J. *Proc. R. Ir. Acad.* **88B**, 39–43 (1988).
3. Nash, L.K. *J. chem. Ed.* **61**, 981–984 (1984).
4. Everett, D.H. *J. chem. Soc.* 2566–2573 (1960).

LOW AND HIGH ESTIMATES OF DIVERGENCE ATTAINABLE BY VARIOUS POPULATION SIZES

Population size	Mutation rate per nucleotide per year		
	2×10^{-8}	10^{-8}	7×10^{-9}
	Range of divergences		
2,000	0.0016–0.0064		
3,000	0.0024–0.0096	0.0012–0.0048	
5,000	0.004–0.016	0.002–0.008	0.0014–0.0056
8,000		0.0032–0.0512	0.00224–0.0179
10,000			0.0028–0.0112
20,000			0.0056–0.0224
Range of population sizes	2,000–5,000	3,000–8,000	5,000–20,000

The three estimates above are based on the values of mutation rate usually cited in the literature^{3,8}. The range of population sizes is the range of sizes of the female population that could have given rise to a divergence of 0.57% for each mutation rate.

genomes of about 0.3 to 0.57% per nucleotide^{2,3} implies a small average population size of about 5,000 females. Were the average population size any larger, the accumulated divergence would be greater and the “coalescence time”⁴, or time back to a single mitochondrial-type, would be correspondingly longer.

If the amount of divergence is known accurately, so is the coalescence time^{3,6}. There is much more uncertainty about the average size of the population leading to a given divergence⁷. The table shows that the population could easily have been smaller than Maynard Smith suggests. Such small sizes would have to be maintained for thousands of generations, with an attendant high risk of extinction.

A second and more likely possibility is that there was a pronounced bottleneck in numbers at about the coalescence time. After this, the human population could

and then proceeding to the blunt end along the promenade deck. It may be that the observed human mitochondrial coalescence time is more than just a statistical accident, and signals a dramatic event in the history of our species.

CHRISTOPHER WILLS

*Department of Biology,
University of California, San Diego,
La Jolla,
California 92093, USA*

1. Maynard Smith, J. *Nature* **344**, 591–2 (1990).
2. Brown, W. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3605–3609 (1980).
3. Cann, R. L., Stoneking, M. & Wilson, A. C. *Nature* **325**, 31–36 (1987).
4. Kingman, J. F. C. *J. appl. Prob.* **19A**, 27–43 (1982).
5. Nei, M. & Li, W.-H. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5269–5273 (1979).
6. Wills, C. in *Molecular Genetic Medicine* (ed. Friedmann, T.) (Academic, New York, in the press).
7. Avise, J. C., Neigel, J. E. & Arnold, J. J. *molec. Evol.* **20**, 99–105 (1984).
8. Nei, M. in *Population Genetics and Molecular Evolution* (eds Aoki, K. & Ohta, T.) 41–64 (Springer, Berlin, 1985).

Trouton's rule

SIR—John Maddox¹ pokes gentle fun at chemists, stating that they “are masters at elevating empirical generalizations into natural laws”. He cites Trouton's rule as a classical example of this practice. However, I have recently published a simple statistical mechanical rationale of this rule², after one expert in the field admitted³ that a satisfactory explanation had never been given, such as that achieved by Einstein for the parallel Dulong and Petit's law of heat capacities of solids. My explanation illuminates the continuity in partition functions in describing changes of state from ideal molecular solids to corresponding ideal liquids to ideal gases. Not only does it reveal the connection with the Einstein heat-capacity theory, but it also shows that the original Eyring transition-state

theory for kinetics is immediately relevant in handling the change from liquid to vapour.

The best analytical form⁴ of Trouton's rule is given by the expression $\Delta S_b^\ddagger = R(4 + \ln T_b)$, where $\Delta S_b^\ddagger$ is the molar entropy involved, and T_b is the boiling point. This equation is very readily derived using the following molar partition functions for ideal systems.

1. Solid (Einstein)
 $(kT/h\nu)^{3N}$
2. Liquid (cell theory)
 $(kT/h\nu)^{3N} \times e^N$
3. Gas
 $(2\pi mkT/h^2)^{3N/2} \times V^N/N!$
4. Liquid/gas transition state
 $(2\pi mkT/h^2)^{3N/2} \times (N\delta^3)^N/N!$

More pebbles

SIR—Lorang and Komar pointed out in Scientific Correspondence¹ that abrasion forces unevenly affect the surface of sea-shore pebbles. If the force comes evenly from any direction, what will be the resultant shape of the pebbles?

I have some ‘experimental’ evidence, obtained from my hobby of china-making. When kaolin is manufactured, local rocks, crushed in a ‘ball mill’, are transformed into clay materials. Rocks with enough water are agitated with hard ‘balls’, which are natural, round-shaped silicated rocks (quartzite). Abrasion gradually acts on the balls.

I noticed that well-used rock balls had three main configurations: flattened ovoid, ovoid and spherical, in the ratio 70, 10 and 20 per cent, respectively. Artificial hard ‘balls’, made from alumina or magnetic iron, are often used instead of the natural balls. These artificial balls keep their original spherical shape even after repetitive usage, although they decrease in size. Thus, I believe that the original structure of the rock material is important for pebble shape, as pointed out by Ashcroft².

TOHRU YAZAWA

*Department of Biology,
Tokyo Metropolitan University,
2-1-1 Fukazawa, Setagaya-ku,
Tokyo 158, Japan*

1. Lorang, M. S. & Komar, P. D. *Nature* **347**, 433 (1990).
2. Ashcroft, W. *Nature* **346**, 227 (1990).

Machine dilemma

Igor Aleksander

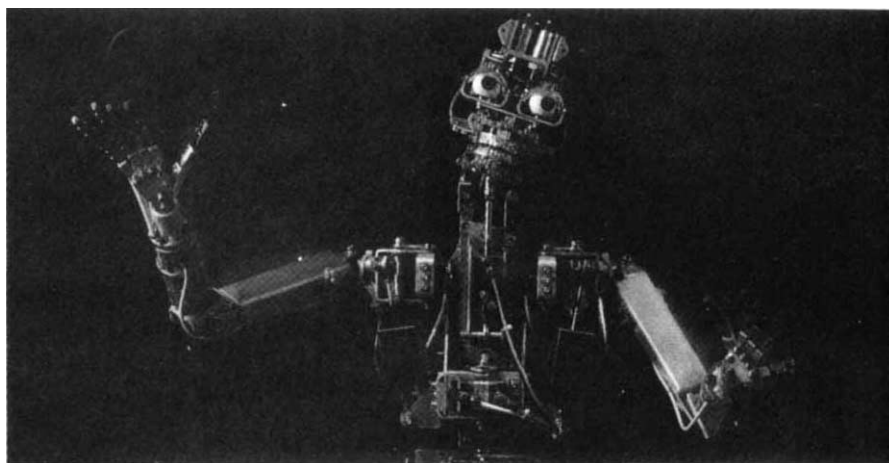
The Age of Intelligent Machines. By Raymond Kurzweil. MIT Press: 1990. Pp. 565. £19.95 (£29.95 after 31 March 1991), \$29.95 (\$39.95 after 31 December 1990).

FOURTEEN-year-old Margaret Litvin, in her seventh-grade essay on "living in the age of intelligent machines", has written "... All right, suppose that we have an artificial brain that is in all senses equivalent to ours ... Why should anyone be attracted to this repulsive idea? ... Besides, I think most reasonable people would agree with me that it won't work anyway." Raymond Kurzweil has included this extract in his collection of the thoughts of those working in artificial intelligence (AI) who evidently have been attracted to the repulsive idea. Unwittingly, he has pointed to a dilemma that sits squarely on the shoulders of anyone working in the field that his book celebrates. Artificial intelligence is now more than 40 years old (although roots may be traced further back). It is concerned with doing things on computers that, if done by humans, would be thought to require intelligence. A massive amount of scientific, intellectual and financial investment appears to have resulted in very little, either as a pointer to the way our brains work (classical AI researchers are quite adamant about not being interested in how the brain works — only in apeing what it does) or improvements in the usefulness of computers. So the dilemma lies in the fact that while doing research in AI may be enormously stimulating for those doing it, justification for it and predictions of what it would achieve have mostly been ill-founded.

Kurzweil is living proof of the nature of this dislocation. He is hailed as one of the leading inventors in the United States for having produced a reading machine that scans written text and turns it into synthesized speech. This is of considerable benefit to the blind, and may be found doing useful work in many libraries and schools. But, such a useful machine has been produced with virtually no appeal to the work of AI researchers — it is based on a high degree of engineering skill, exploitation of what computers do well anyway, and Kurzweil's unbounded entrepreneurship. If machines have proved to be useful for a large number of people it is in the areas of spreadsheets, word processors and databases and not in trying to be like humans.

As Kurzweil's invention of the reading machine was a heroic act, so is his book, *The Age of Intelligent Machines*. It is large, heavy, contains a mixture of contributed and self-authored passages, photographs of self-satisfied looking AI geniuses,

and sports a mixture of writing styles ranging from short essays to platonic dialogues with highly illustrated explanatory passages somewhere in between. Virtually the whole of civilized thought is seen as the basis of artificial intelligence: in the first part of the book the reader is



Robot Joral — a repulsive imitation of man? Photo: Len Jones.

treated to potted versions of the thoughts of philosophers from Plato to Wittgenstein. This is supplemented with essays by modern AI "philosophers" among whom are Daniel Dennet ("... preoccupation with the Turing test is regrettable ..."), Sherry Turkle ("... [have] people always thought like machines? ...") and Douglas Hofstadter ("... I am an unabashed pusher of the Turing test. ..."). The essays are more interesting than the potted philosophy, with something in them both for those who know the arguments and those who do not.

It is hard to separate the mathematical basis for artificial intelligence from that of computer science in general. But this is attempted in a short chapter which is largely on logic. Kurzweil is much impressed by the generality of the "NOR" logical function and the fact that any digital machine could be constructed out of a sea of NOR logical "gates". This is fine, but displays a curious gap in what is otherwise quite a thorough coverage. There is no reference to the pioneering work of John von Neumann on the general logical theory of automata and self-reproduction in cellular automata. These make for a far more interesting way of approaching the "universality" problem than through the NOR gate. Mathematical roots are followed by technological roots, and it is in this area that the book comes to life. The history of both the mechanical side and

the electronic side of computers (rather than AI) is well illustrated and appealingly written.

The second half of the book concentrates on the more recent aspects of AI — knowledge-based systems. Within this, Kurzweil's appreciation of the problems of recognizing visual and auditory patterns leads to lucid descriptions of some simple ways of solving these. After all, such methods have led to the success of his own machines. But such is the snobbery of researchers in rule-based systems, that I doubt they would accept hard-nosed programming methods such as "template matching" as being important

to AI. Also, those who are looking for a description of current debate in neural networks will find only a passing reference to the subject. This may be ironic as, not only does this field provide some hope of achieving some unaccomplished targets in AI, but also puts a bomb under the philosophical debate, centred as it is on the relevance of needing to represent intelligence by a set of computable rules. Neural systems, like living systems, learn and adapt, leading to an emergent intelligence within the machine rather than one that is entirely the responsibility of a programmer who enjoys playing God.

Despite the limitations, this is an adventurous book. Both novices and experts will find in it something new and educational. For me, for example, the high spot was the discovery of Margaret Litvin's essay, which does more to prove the supremacy of the wisdom of people than anything a computer is ever likely to accomplish. To those who argue that some day computers will act as if they understand, think, love and hate, Margaret says: "Why can't scientists enjoy people, who really do these things?" Is the AI dilemma a result of scientific envy of naturally produced intelligent beings, who do things so well? □

Igor Aleksander is in the Department of Electrical Engineering, Imperial College of Science, Technology and Medicine, Exhibition Road, London SW7 2BT, UK.

Escalators to glory

Simon Conway Morris

Evolutionary Trends. Edited by Kenneth J. McNamara. *Belhaven*: 1990. Pp. 368. £45.00. University of Arizona Press, \$24.95.

THE history of life has been described as a shimmering tapestry. Just as many of the most interesting patterns in a tapestry become clear when viewed across the room, so perhaps some of the patterns of evolution can only be seen through palaeontological spectacles. With so many evolutionary trends part of our folklore, it is a brave editor who summons the evolutionary hierarchs. That McNamara has been largely, if not quite, successful will be apparent not only from the range of expertise (indeed one feels the authors will learn as much from each other as any reader), but more significantly because the reiteration of common themes makes this book an unexpected treasure-trove for the biologically curious.

One major question is what relationship lies between the generation of trends and environmental factors. In the case of marine bivalves (documented by Miller), it appears that it is the environment in which these creatures find themselves, rather than net changes, that drive the evolutionary trend. In a related vein McNamara illustrates how evolutionary trends in infaunal echinoids arise by progressive invasion of fine-grained sediments, often in deeper water. Discussing mammals, however, Janis and Damuth are even more committed to the role of the environment, writing that "more or less unidirectional climatic change . . . has driven much of mammalian evolution and diversification". Others are more muted. For example, Dommergues writing on ammonites finds no clear connection between "morphological change and possible environmental constraints".

The trends themselves may fall fairly and squarely into macroevolutionary patterns, but what of the processes? Here this book is somewhat indecisive. Gould's opening chapter is cast into a thoroughgoing macroevolutionary framework, but elsewhere the issues are evaded. There are some perfunctory nods to punctuated equilibria, but more effort is expended by several authors in balancing processes involved with anagenesis and cladogenesis. A more constant theme is the role of heterochrony. If Gould put heterochrony back on the map, then it is largely thanks to McNamara that we have a recognizable geography. Scarcely a chapter fails to invoke some sort of heterochrony, but as to actual mechanisms that change timing

of development we are largely ignorant. But even if pattern rather than process is understood, proper comprehension of evolutionary trends allows sensible projection into the future. So on the last page McNamara unfolds the likely history of a group of infaunal echinoids. Readers of *Nature* in a mere two million years will be raiding their library stacks to check this prediction.

It is generally accepted that metazoans occupy only a small portion of the potentially available morphospace (however defined). A major pre-occupation of this book is their trajectories through time. Cited by almost all the contributors is Gould's observation that in many cases trends are a result of no more than changes in variance through time. But also relevant is how trends are constrained. One aspect relates to the boundaries (or reflectors) of morphological possibility. This is explored by McKinney who adopts a nomothetic approach, with emphasis on serial correlation of trends. The topic is

touched on by others, principally by Miller in his discussion of on-shore-off-shore trends.

Morphological trajectories presuppose a destination, and a striking aspect of this volume is the emphasis on convergence. So commonplace is this evolutionary phenomenon that comment might seem superfluous. This would be a mistake, and the reiterated examples of convergence in groups such as mammals (Janis and Damuth), ammonites (Dommergues), bivalves (Miller) or echinoids (McNamara) show that any discussion of trends must involve analysis of why the clouds of morphology so often condense at the same spots. This is no trivial observation, because it undermines those who suppose that if life's tapestry was unwound and rewoven the pattern would be very different.

Simon Conway Morris is in the Department of Earth Sciences, University of Cambridge, Downing Street, Cambridge CB2 3EQ, UK.

Crystals on the map

Moreton Moore

Historical Atlas of Crystallography. Edited by J. Lima-de-Faria. *Kluwer*: 1990. Pp. 158. Dfl 69, £20, \$20.

"He who sees things grow from the beginning will have the best view of them" (Aristotle).

Crystallography has been an international science from the earliest times. The atlas contains some maps of Europe to show where the principal crystallographers worked in the sixteenth to the nineteenth centuries as well as a map of the world to include twentieth-century pioneers from North America, India and Japan. This is also a historical atlas in

which the coordinate of time is represented on a linear scale and important events marked in their correct positions on such "time maps". Thus one can see at a glance when the periods of greatest achievement took place, in a graphic and memorable display which will appeal to practitioners of this visual science. The drama is brought alive by portraits of the "actors" and by reproductions of the title pages of their early works.

The four main branches of crystallography — geometrical, physical, chemical and crystal structure determination — are each treated historically in the chapters which follow. Probably the first textbook on mineralogy was *De Natura Fossilium* by Agricola (alias Georg Bauer), published in Basel in 1546. In Copenhagen in 1669, the first important experimental discovery in crystallography was made by Bartholinus (Bartholin) when he observed the double refraction of light

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Crystallography triumphant — X-ray diffraction image of DNA.

passing through Iceland spar (calcite); and in the same year the law of constancy of interfacial angles was proposed by Steno (Steenen) from his measurements of quartz crystals. The first definition of crystallography as a new science appeared in Romé de l'Isle's *Essai*, published in Paris in 1772; and he taught the first course on crystallography there from about 1783. An example of international activity is the derivation (from 1880 to 1884) of the 230 crystallographic space groups by Fedorov (St Petersburg), Schoenflies (Göttingen) and Barlow (London). Often what had seemed to be only of academic interest has later become of practical importance.

Crystallography is also interdisciplinary, making major contributions to biology, chemistry, metallurgy, mineralogy and physics. It has developed rapidly since the discovery in 1912 of the diffraction of X rays by crystals. In Munich at that time, Wilhelm Röntgen was Head of the Institute for Experimental Physics, Arnold Sommerfeld was Head of the Institute for Theoretical Physics and Paul von Groth was Head of the Institute for Mineralogy and Crystallography. These three intellectual giants, together with their assistants and graduate students, customarily met after lunch for coffee and cakes at the Café Lutz in the Hofgarten. Into this creative atmosphere came Max von Laue, Paul Ewald (who was writing his doctoral thesis on diffraction), Walter Friedrich (Sommerfeld's assistant) and Paul Knipping (a graduate student of Röntgen). As is well known, Friedrich and Knipping performed the first X-ray diffraction experiment under von Laue's general supervision; but the location of the atoms in the crystal structures was performed in England by the father-and-son team of William Henry Bragg and William Lawrence Bragg. Thus the subject of crystal structure analysis began (in 1913) and over the years this has contributed immensely to our understanding of minerals, metals, proteins, enzymes and viruses.

Throughout the history of crystallography, success in instrumentation has gone hand-in-hand with developments of relevant theory; and in the last fifty years, advances have been made with the help of ever-increasing computational power. The *Historical Atlas of Crystallography* tells the stories of these developments, through the time maps and the individual chapters; and by copious references to the sources. The *Atlas* will be of interest not only to crystallographers but also to historians of chemistry, mineralogy and physics. Indeed, it will surely become a paradigm on which future texts will be modelled. □

Moreton Moore is in the Department of Physics, Royal Holloway and Bedford New College, University of London, Egham, Surrey, TW20 OEX, UK.

Cultural exchange

Robert M. Macnab

Physiology of the Bacterial Cell: A Molecular Approach. By Frederick C. Neidhart, John L. Ingraham and Moselio Schaechter. 1990. Pp. 506. United States, Sinauer: \$43.95; Britain, Freeman £34.95.

GROWTH, although certainly a fundamental activity, is not the only one bacteria undertake. *Physiology of the Bacterial Cell*, which has evolved from a previous book titled *Growth of the Bacterial Cell* (Sinauer, 1983), wisely acknowledges this fact and includes treatment of such topics as chemotaxis, cellular differentiation and ecology, as well as the previous thorough treatment of cell composition, structure, assembly, metabolism and growth.

The authors take *Escherichia coli* as their "reference cell", and for data and references draw heavily from the now standard two-volume text *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* by J. L. Ingraham *et al.* (American Society for Microbiology: 1987). But they do make a real effort on each topic to give at least some indication of how other species go about their business. Also, the book does not claim to provide a broad description across 'the microbial world' (a description that is already available in an existing book under that title by Ingraham and co-authors).

Intended mainly for upper undergraduate or graduate students, but also useful as a reference source for others, *Physiology of the Bacterial Cell* is written by three masters in the art of communicating in a fashion that is simple and lucid without being condescending. It is informative and up-to-date, providing an appropriate level of detail for a survey book and pointing the reader to more detailed treatments elsewhere. The authors make extensive and effective use of numerical values and calculations concerning all aspects of the cell's existence, such as the number of molecules of various sorts that the cell contains, or the error rates and costs of correction associated with the successive stages of gene expression. The book is well illustrated, striking a happy balance between reaction schemes, skilfully drawn cartoons of processes such as membrane assembly, and light and electron micrographs.

The reader is first introduced to the bacterial cell from the point of view of its chemical composition, and is then given a survey of its ultrastructure. The treatment of metabolism, by the authors' own admission, follows an unconventional order,

starting with synthesis and assembly of macromolecules (DNA, proteins) and macromolecular complexes (cell membranes, flagella), then continuing with the synthesis of small molecules, and finishing with what is usually the starting point, the main catabolic ("fueling") pathways of the cell; this order presented no problems for me, and would probably present none even to an inexperienced reader. Acquisition of food, both in the sense of seeking it out (chemotaxis) and taking it up (active transport) are given treatment in a separate chapter.

The remainder of the book deals principally with growth, regulation of metabolic activity (at the level of enzyme activity and gene expression) and genetic adaptation. Here, even though each individual chapter reads clearly enough, the overall organization could be improved. For example, a chapter on growth of cells and populations is separated from chapters on the cell cycle and on growth rate by several chapters on unrelated material such as genetic adaptation.

Overall, this is a commendable book that should be a pleasure to read for the student interested in getting a broad introduction to bacterial physiology. □

Robert M. Macnab is in the Department of Molecular Biophysics and Biochemistry, P.O. Box 6666, 260 Whitney Avenue, New Haven, Connecticut 06511, USA.

Copies of articles from this publication are now available from the UMI Article Clearinghouse.

Yes! I would like to know more about UMI Article Clearinghouse. I am interested in electronic ordering through the following system(s):

- ☐ DIALOG/Dialorder ☐ ITT Dialcom
☐ OnTyme ☐ OCLC ILL Subsystem
☐ Other (please specify) _____
☐ I am interested in sending my order by mail.
☐ Please send me your current catalog and user instructions for the system(s) I checked above.

Name _____
 Title _____
 Institution/Company _____
 Department _____
 Address _____
 City _____ State _____ Zip _____
 Phone (____) _____

UMI Article Clearinghouse

Mail to: University Microfilms International
 300 North Zeeb Road, Box 91 Ann Arbor, MI 48106

At home on the farm

D. M. Broom

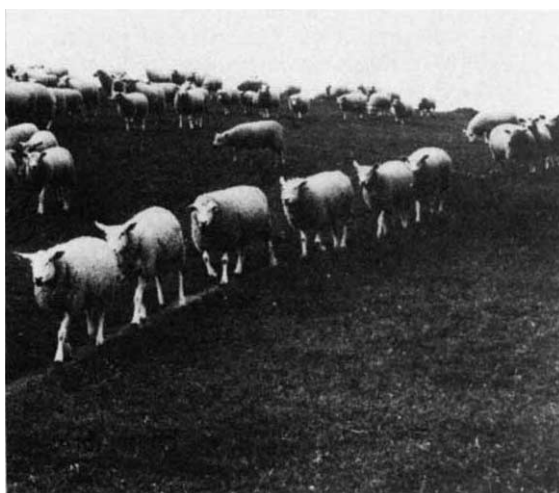
Domestication: The Decline of Environmental Appreciation. By Helmut Hemmer. Cambridge University Press: 1990. Pp. 208. £45, \$69.50.

DOES domestication involve "a special kind of regressive evolution" and is the behaviour of domestic animals "weaker and less determined by environmental factors than that of wild animals"? Helmut Hemmer presents these arguments in this new edition of a book published in German in 1983 and adds a subtitle to the book to emphasize his opinion. It is widely accepted that one of the genetic differences between wild and domesticated forms of a species is a smaller propensity on the part of domestic animals to show adrenal responses to man as a source of danger. This difference has wide ranging consequences for the behaviour which we see shown by wild and domesticated strains, even when these have been reared in similar ways. But Hemmer's idea that perceptual abilities in general are less good in domesticated animals is not clearly supported by the evidence available in this book or elsewhere. The judgement about the direction of evolution is also questionable because domestic animals are among the most successful species of their classes in numerical terms and may be thought of as efficient exploiters of an ecological niche.

But Hemmer's thesis is intriguing and cannot be lightly dismissed. Comparative data on brain size shows that domestic forms have relatively small brains. There could have been some reduction in brain size in domestic forms, or man could have selected the forms with smaller brains, or domestication may have taken place in the warmer parts of the species' ranges where smaller-brained forms occur. A further proposal is that domestic animals can live in very large herds because they "are scarcely aware of food competition and the aggression it provokes". Shades of the captive rabbits described in Richard Adams' *Watership Down*, but evidence concerning the behaviour of domestic animals conflicts with this view. The complex social organization and impressive learning ability of several species of farm and companion animals makes the idea of overall degeneracy untenable.

Half of the book is a fascinating descriptive account of domesticated animals with especial emphasis on the evidence

for the derivation of the various species. For example, the domestic dog is convincingly argued to be descended from the southern wolf *Canis lupus pallipes* of Arabia and Northern India because of smaller canassial teeth, smaller coronoid processes on the lower jaw, different facial expressions, shorter vocalizations and connected pads on the third and fourth toes. In accounts of attempts to mimic foxes by Belyaev and Trut is quite well known, but a detailed account is also provided of research carried out by Hemmer's own group on the selection of fallow deer. The new forms produced are less readily alarmed, easier to handle,



Falling into line — the most domesticated animal?

faster in growth and more likely to be sexually active in the second year of life than unselected forms. Coat colour seems to have been a major factor in the selection of this and several other species mentioned. The author makes the provocative suggestion that coat colour is related to reactivity because pigments and catecholamine transmitters share a common biochemical synthetic pathway.

The ideas about the general characteristics of domesticated animals are put forward with frequent reference to stress. Contrary to the general idea that stress is an environmental effect which has adverse consequences for an animal, Hemmer defines it as "a mathematical function of the stimuli impinging on an individual and the information which it acquires by processing them". The definition is so wide ranging as to be almost meaningless but is accompanied by some perceptive comments on how domestic and wild animals are affected by difficult environmental conditions. Several of the concepts presented will stimulate controversy and research and all those interested in evolution or economically important animals will enjoy reading the book. □

D. M. Broom is in the Department of Clinical Medicine, University of Cambridge, Maddingley Road, Cambridge CB3 0ES, UK.

New environmental books

Along with increasing concern about the state of the planet has come an ever-expanding list of books on environmental and related issues. Needless to say, it would be impossible to review all such books within the pages of *Nature*, but what follows is a sample of some of the new books on offer.

■ *One Earth, One Future: Our Changing Global Environment* explains the many and varied ways in which the activity of mankind has so often led to detrimental changes in the global environment. In the first section of the book entitled 'The Earth as a System' coauthors Cheryl Simon Silver and Ruth DeFries discuss the interdependence of major ecological systems such as the atmosphere and the oceans in the context of the Earth's geological past. They go on in the second section 'The Faces of Global Environmental Change' to cover such problems as global warming, ozone depletion, acid deposition and biodiversity. The book is published by the National Academy of Sciences, price \$14.95 (\$18 overseas).

■ Global climate change and the greenhouse effect have attracted a good deal of attention recently. Two books of particular interest are *Climate Change and World Agriculture* by Martin Parry (Earthscan: 1990; pbk price £9.95, \$19.40), and *Climate Change and Plant Genetic Resources* edited by M. Jackson, B. V. Ford-Lloyd and M. L. Parry, both of which assess the likely impact of global change on our ability to feed the world's hungry. Published by Belhaven; price £27.50, \$49.00.

■ Several recent books have attempted to explore possible solutions for thorny environmental problems. Of special note are the following books, each of which is in its way highly critical of present-day policy makers: *Making Peace with the Planet* by Barry Commoner published by Victor Gollancz Limited, price £16.95 (also published in the United States by Pantheon Books, price \$19.95); *Barriers to a Better Environment: What Stops Us Solving Environmental Problems?* by S. T. Trudgill and published by Belhaven — price £25.00 (also published in the United States by Columbia University Press, price \$54.50).

■ The relationship between the environment and human health is addressed by two new books. In *Health and the Global Environment* Ross Hume Hall argues that by failing to take into account environmental issues when designing policies on a wide range of issues the powers that be are endangering human health in the long term. Published by Polity Press, price £29.50, \$44.95. *Environmental Health* edited by J. Rose assesses the impacts of pollutants on health. The book is published by Gordon and Breach price £45, \$90.

■ In *A Year in the Greenhouse: An Environmental Diary* published by Victor Gollancz, environmental writer John Elkington sets out to give an insider's view of the Green movement at a time when world leaders are discovering the electoral benefits of environmental friendliness, price £16.95.

Correlation of solar neutrino modulation with solar cycle variation in p-mode acoustic spectra

Lawrence M. Krauss

Center for Theoretical Physics and Department of Astronomy, Sloane Laboratory, Yale University, New Haven, Connecticut 06511, USA

A tantalizing correlation (at the 0.97–0.999 level) between reported variations in low-order acoustic solar oscillations over the solar cycle from 1977 to 1988 and the event rate in the ^{37}Cl solar-neutrino detector over the same period is demonstrated. Such a correlation could provide more detailed evidence on how modulations of the solar neutrino flux might result from variations in the Sun.

SEARCHING for correlations between any *a priori* unrelated phenomena can be both subtle and dangerous. The solar cycle—manifested by the number of sunspots appearing on the Sun—has been claimed to be correlated to almost everything under the Sun, from the stock market to the solar-neutrino detection rate. Are any of the claimed correlations statistically significant? Recently, Bahcall and Press¹ have made the case that the ^{37}Ar production rate in the University of California (Davis) ^{37}Cl detector over the past 20 years displays a statistically significant modulation over the solar cycle. (As this work was completed I received a preprint² in which sunspot-neutrino correlations are also examined.) They also made the interesting suggestion, based on Fourier analysis of the data, that this modulation might not have the same shape as the sunspot cycle. They postulated that the correlation may not be directly related to sunspots, but rather to some other feature coupled to the solar cycle. This is not unreasonable, because only a model in which the neutrino magnetic moment is near its maximum allowed value predicts such an (anti) correlation of neutrinos with sunspots^{3,4}. These authors, however, did not find a background activity that correlated well with the observed neutrino signal.

An analysis of 11 years of continuous monitoring of solar p-mode acoustic spectra has just been reported⁵. In the period 1977–88, the frequencies of the lowest order ($l \leq 2$) modes varied with a peak-to-peak amplitude of $0.46 \pm 0.06 \mu\text{Hz}$ in a manner that seemed to correlate with sunspot number. The amplitude of the shift suggests a change in travel time through the Sun of the order of one second. This time is strongly weighted by travel near the surface, where the speed of sound is slowest, and the p-mode eigenfunctions are dominant. Nevertheless, the signal from low- l modes is coupled more directly to variations beneath the surface than are sunspots. Variations in temperature, internal magnetic field or solar dimension could affect it, so that a correlation with neutrino detection could provide new information to help understand modulations in the neutrino event rate. This was the motivation for my examination of the correlation between solar oscillations, sunspots and neutrino event rates. Besides examining linear correlations, the shapes of the neutrino modulation and acoustic oscillation shift were examined by a weighted Fourier periodogram analysis. The p-mode oscillation shift seems to decrease in period with increasing mode number, consistent with an apparent improvement of the linear correlation of the neutrino data with increasing mode number. These results support a possible coupling between solar-neutrino event

rates and variations in the solar cycle, and may be relevant for any explanation of the deficiency of the neutrino event rate.

The data, a first pass

I first coordinate the data on a baseline appropriate to the observations of the solar oscillations, which happen to have been made over the 11-year period in which the neutrino–sunspot correlation is strongest¹. Not all of the neutrino–sunspot data, however, can be compared with the oscillation data. The power spectra from which the observed shifts are inferred were composed of observations over 1–2 months in each year up to 1988. For each l value there are therefore only 13 distinct data points, a fact that has implications for the strength of the conclusions one can draw. These are shown, for $l=0, 1$ and 2 , in Fig. 1a. Error bars are quoted standard deviations. The mode-averaged shift for each time is shown in Fig. 1b. For comparison, I isolated the sunspot and neutrino data obtained during the month or months of oscillation data. This was easiest for the sunspot data, which was completely and evenly sampled in time. (The neutrino and sunspot data used here were taken

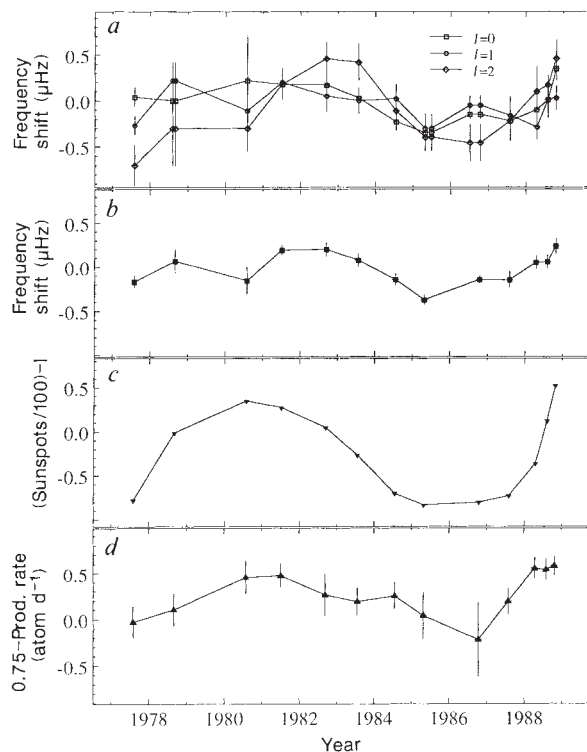


FIG. 1 Solar-cycle modulations of solar acoustic oscillations (the separate $l=0, 1, 2$ modes are shown in *a* with 1-standard-deviation errors and in *b* the statistical average of the mode shifts at each time is shown), *c*, sunspots and *d*, the event rate in the ^{37}Cl detector (error shown is symmetrized value of upper error bar from data). The data in *b* and *c* has been reduced to the time baseline of the acoustic-oscillation measurements.

from refs 1 and 6. In particular the neutrino data used here include a subset of those used in the analysis in ref. 1.) The monthly average, or if two months were spanned, the average of the monthly averages was used. The result is shown in Fig. 1c. Errors were not estimated for this data.

A similar procedure was used for the neutrino data⁶. All observing periods for the oscillation data happen to overlap with at least one neutrino observing period. When more than one neutrino point overlapped with the oscillation data period, a weighted average was used (when the error bars were asymmetric because the lower limit was below zero, the upper error bar was used to determine statistical averages). There is a lot of jitter in the raw neutrino data and the event rate at one time can differ significantly from the mean value obtained at adjacent times. To reduce the likelihood of biasing due to a chance excursion from the local mean, I averaged the neutrino data for the period(s) corresponding to the oscillation data with the data points in adjacent time bins. This considerably reduces the noise, and produces a fairer sample of the actual temporal behaviour of the neutrino data. (Because of the gap in the neutrino data in 1985–86, the data used during this time involves only the bins during and immediately before first of the two oscillation data periods in 1985, and during and immediately after the latter of the two in 1986. Hence the larger error bars in these two cases.) The result is shown in Fig. 1d, where the sign of the data is reversed and is shifted by a constant so that an anticorrelation with the other data is displayed on the same scale as a correlation.

A correlation between the data sets is apparent. To determine its nature and significance, I used statistical tests, including those described in ref. 1. These shed light on the question of correlations and on how incorporating the quoted errors may affect the conclusions.

Figure 2a shows the neutrino data (with the sign reversed) plotted against the oscillation shifts, and Fig. 2b shows the neutrino data plotted against sunspot number. An anticorrelation should appear as a rise from the lower left to the upper right. A weighted linear least-squares fit to the data, treating the neutrino event rate as the dependent variable, yields χ^2 values that should be exceeded only 40% (Fig. 2a) and 30% (Fig. 2b) of the time, and provides no strong evidence for linear correlations. The same is true for the oscillation–sunspot data (not shown). But such a correlation may be beyond the sensitivity of the χ^2 test. One can look for evidence against any correlation by performing a χ^2 test on the data fitted to a constant value and comparing this to the linear fit. The constant fit is poor, with a χ^2 value that is likely to be exceeded <1% of the time. The likelihood that another linear term is needed can be estimated using the *F*-test, in which ratios of χ^2 values for the two fits are compared. For the neutrino–oscillation data the F_χ statistic is 7.45, for the neutrino–sunspot data $F_\chi = 11.5$, and for the oscillation–sunspot data $F_\chi = 37.3$, yielding 98%, 99.3% and 99.95% confidence levels, respectively, that another linear parameter is needed to fit the data. (The *F*-test described here

was also performed in ref. 2 to examine the neutrino–sunspot correlation (on a strongly averaged version of the data from 1970–89).

Although interesting, these results have limitations. The distributions are not known *a priori* to be gaussian. More important, they weight the uncertainty in the neutrino data, but not the uncertainty in the other data. If I reverse the dependent and independent variables and perform new fits to the oscillation–neutrino data, the linear fit χ^2 jumps from 15.4 to 49.2, and from 25.8 to 91.6 for the constant fit, owing to the noisiness of the oscillation data. This limits how good any correlation with the neutrino data can be: the confidence attached to a correlation should not be better than the confidence in the data.

Refined correlation probes

A more widely used test for linear correlations between data than the simple χ^2 test uses the linear correlation coefficient, r (refs 7, 8). For N pairs of data (x_i, y_i) , all of which have the same variance,

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}} \quad (1)$$

where $\bar{x}$ and $\bar{y}$ are the mean values. If the distribution for x and y is approximately normal, the probability that the magnitude of r should exceed its observed value under the null hypothesis is determined using the statistic $t = r[(N-2)/(1-r^2)]^{1/2}$, which follows a student's t distribution with $\nu = N-2$ degrees of freedom, even for N not large⁸.

When both sets of data points have individual uncertainties σ_i that are not uniform, equation (1) must be refined. One should weight individual data points by $1/\sigma_i^2$ to give weighted mean values $\bar{x}$ and $\bar{y}$. One can also weight each data point in the products and sums in equation (1) in a similar manner, giving

$$r = \frac{N \sum_i \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x^2 \sigma_y^2}}{\sqrt{\sum_i \frac{1}{\sigma_x^2} \sum_i \frac{1}{\sigma_y^2}} \sqrt{\sum_i \frac{(x_i - \bar{x})^2}{\sigma_x^2} \sum_i \frac{(y_i - \bar{y})^2}{\sigma_y^2}}} \quad (2)$$

This is a generalization of the weighting procedure in ref. 7 for the case where both data sets have varying uncertainties. (I have not seen this expression given explicitly in the literature.)

To probe both linear correlations and the effect of incorporating varying uncertainties, I calculated both versions of r . Table 1 presents the (two-sided) confidence limits (CL) for these as well as for the non-parametric tests described below. An inferred anticorrelation of the averaged oscillation data and the neutrino data is good (CL = 97%), although not indisputable. The correlation between neutrinos and sunspots is better (CL = 99.2%), and is strongest between oscillation shifts and sunspots (CL = 99.98%). It seems plausible that the difference between the strength of the neutrino–oscillation and neutrino–sunspot anticorrelation reflects the noisiness of the oscillation data compared to the sunspot data. Given the uncertainties in the neutrino and oscillation data, the strength of the anticorrelation is impressive. If I consider all 39 oscillation shifts together in a fit to neutrino data (as the correlation between oscillations and sunspots was originally inferred⁵) the formal confidence level increases to 99.9%. Thus the oscillation–neutrino anticorrelation would seem overwhelming, although still less strong than the oscillation–sunspot correlation. This procedure may not be appropriate, however, as the three sets of different mode measurements could be correlated and the relevant neutrino data set contains only 13 independent points.

Even if I do not use all oscillation-shift data in a single fit, it is useful to consider the anticorrelations between the separate l modes and the neutrino data. Here an apparent trend emerges,

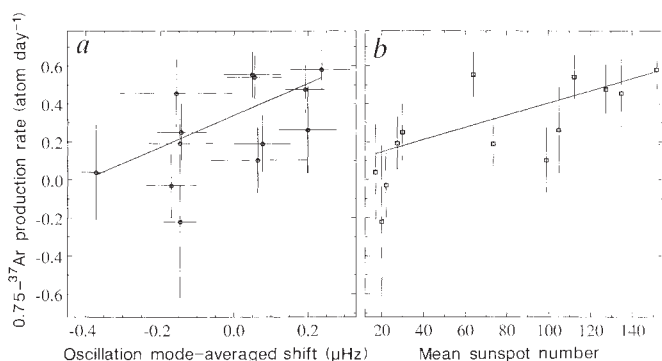


FIG. 2 ³⁷Ar production rate (reversed in sign and shifted by a constant) plotted against *a*, averaged oscillation shift and *b*, sunspot number, as given in Fig. 1. Also shown are the least-squares linear fits.

with increasingly significant anticorrelation for increasing mode number (no such trend is observed in the oscillation-sunspot correlation (not shown)). This is not apparent if the unweighted r coefficient is used, demonstrating how ignoring the errors attached to the data points can alter the results. (The unweighted r is also smaller for the oscillation-neutrino and oscillation-sunspot data sets, and increased for the neutrino-sunspot combination.) To explore further how the uncertainties affect the r coefficient, I examined the correlation of the (noisier) unaveraged neutrino data with the (averaged) oscillation and sunspot data, and also the change in correlation if a single uncertain neutrino (or oscillation) data point was moved within its 1σ limits. The unaveraged data produced a uniformly lower weighted r correlation, whereas the unweighted r statistic varied haphazardly (with the neutrino-oscillation anticorrelation being more strongly suppressed). Moreover, the weighted statistic was more robust under variations of the data within its uncertainty limits (showing the need for more accurate data before correlation estimates can be improved).

Unless a correlation is known to be strong, the significance derived from parametric fits depends on the distribution from which the data points are sampled. An alternative and potentially more sensitive test of correlations can be made by so-called non-parametric tests, which include the Spearman rank-order coefficient, and Kendall's τ -coefficient^{1,8}. In these, the data points are replaced by their respective integral rank order in each data set, and the correlation of the ranks is investigated, either by a linear correlation coefficient between ranks, or a relative comparison of ranks. This has the advantage that the underlying distribution of ranks is known—it is uniform among the integers. It has the disadvantage, however, that the error bars on each data point are not used.

I applied both tests to the data sets treated earlier, and the results are given in Table 1 (two independent confidence limits for the Spearman test, obtained from the variance of the sum of the squared difference of ranks D and from the t statistic for the rank correlation coefficient, give an internal estimate of its uncertainty). The results are striking in their quantitative and qualitative agreement with the results derived from the unweighted r coefficient. I conclude that here these tests have no greater sensitivity to possible correlations than does the unweighted r test. Moreover they suffer from the same disadvantages. They tend to overestimate somewhat the significance of the neutrino-sunspot anticorrelation, and underestimate the significance of the anticorrelation in the noisier neutrino-oscillation data. This could be relevant to the results of ref. 1, where these tests were used to examine the neutrino-sunspot correlation. Here, however, they add no new information.

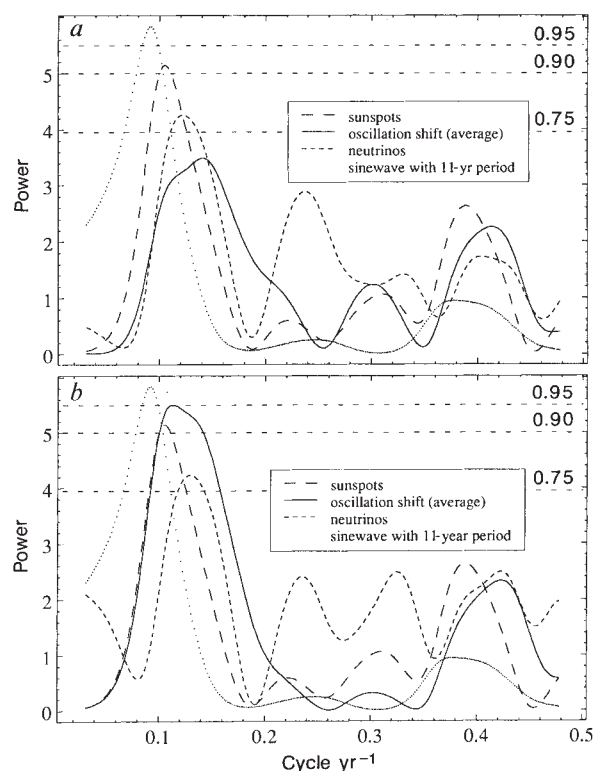


FIG. 3 *a*, Unweighted and *b*, weighted periodograms for the sunspot, averaged oscillation shift and neutrino data in Fig. 1. Approximate confidence levels are also shown. Also shown is the periodogram for a sine-wave with 11-year period sampled over the same intervals as the other data.

One may ask whether the apparent neutrino-oscillation anticorrelation merely results from strong correlation between the oscillation and sunspot data sets, the latter of which is apparently anticorrelated with the neutrino data. To explore this further, I performed a Fourier periodogram analysis for unevenly sampled data⁹, which in ref. 1 gave indications that the shape of the modulations in neutrinos and sunspots were different. I refined the procedure to account for varying uncertainties in the data by weighting the mean, the variance and all sums over the data by the individual σ_i , as in equation (2). Figure 3*a, b* shows the normalized and, respectively, unweighted and weighted periodograms for the neutrino, sunspot and oscillation-shift data, as well as the periodogram for an equally sampled sine curve with an 11-year period.

TABLE 1 Statistical confidence limits for correlations

Data	Linear correlation		Spearman		Kendall
	weighted	unweighted	from D variance	from t statistic	
Averaged ($\nu=11$)					
Neutrino-oscillation	0.97	0.955	0.95	0.945	0.932
Neutrino-sunspot	0.992	0.997	0.998	0.993	0.995
Oscillation-sunspot	0.9998	0.997	0.99	0.982	0.995
Full oscillation ($\nu=37$)					
Oscillation-neutrino	0.999	0.9992	0.9995	0.9993	0.9997
Oscillation-sunspot	0.99995	0.9999	0.9999	0.9997	0.99995
Each mode ($\nu=11$)					
Neutrino-oscillation ($l=0$)	0.82	0.93	0.9	0.906	0.888
Neutrino-oscillation ($l=1$)	0.95	0.80	0.82	0.832	0.777
Neutrino-oscillation ($l=2$)	0.975	0.985	0.992	0.988	0.993
Other ($\nu=11$)					
Neutrino (not averaged) oscillation	0.925	0.86	0.82	0.838	0.802
Neutrino (not averaged) sunspot	0.93	0.98	0.985	0.981	0.988
Neutrino (1σ)-oscillation	0.97	0.965	0.96	0.952	0.949
Neutrino-oscillation (1σ)	0.975	0.988	0.996	0.989	0.989

The sunspot data shows a marginally strong Fourier component at or near 10 years (with confidence at peak power of $\sim 90\%$). The periodograms for the neutrino and oscillation data differ for the unweighted and weighted analyses. The results discussed above suggest that the latter might be expected to be more reliable. The neutrino peak power does not alter in significance, but its location, which is always at a shorter period than the sunspot data shifts towards a yet shorter period in the weighted analysis. Its peak value is less significant (confidence = 80%), but is higher than that obtained if the data from the period 1970–89 is used. (A weighted analysis of all the neutrino data (1970–89) suggests that the most significant peak is at ~ 0.34 cycle yr^{-1} . This peak appears as only marginal in the unweighted analysis of ref. 1. Also, the peak at ~ 1 cycle yr^{-1} is somewhat lowered in the weighted analysis compared to that quoted in ref. 1.) The periodogram for the averaged oscillation-shift data is perhaps the most interesting. Here a dramatic change in significance of the peak power is observed when the data is weighted, from $\approx 50\%$ to a more significant 95% confidence level. Nevertheless, although the peak height changes dramatically, and the shape also changes somewhat, in both cases the mode-averaged oscillation peak remains fairly broad. In the weighted case, there is significant power associated with the periods of both the sunspot and neutrino peaks. In both cases its envelope seems to be centred more closely on the neutrino peak, however, suggesting that to the extent that both the neutrino and oscillation-shift data are sinusoidal, they appear to have (from 1977 to 1988) a shorter period than the sunspot data. This result is consistent with the apparent time lag and lack of a flat plateau in the neutrino data with respect to the sunspot data that was remarked on in ref. 1. Such a time lag in the oscillation and neutrino data, at least in 1980–85, also shows up in the data plotted in Fig. 1.

The broad nature of the oscillation peak and the improved linear correlation of neutrinos with increasing mode number

motivate examining periodograms for the separate shifts in each mode. These are displayed in Fig. 4a, b for the unweighted and weighted cases, respectively, on a magnified scale. The significance of the $l = 1$ mode increases dramatically in the weighted case, consistent with the observation of the enhanced linear correlation with neutrinos in this instance. But even more interesting is the progressive shift of power towards shorter wavelengths with increasing mode number. This accounts for the broadness of the averaged shift peak, and probably also for the enhanced linear correlation with neutrinos.

Because none of the peaks have an overwhelming significance (except perhaps for the $l = 2$ peak), one should not read too much into these results. Nevertheless, if the modulation of the neutrino event rate over the solar cycle is real, as statistical arguments seem to suggest, something must be responsible for this modulation. Given that no plausible terrestrial backgrounds have yet been suggested, it is reasonable to assume that something associated with the Sun is causing the modulation. It is then important to probe how neutrinos may be correlated to other observables that are directly coupled to more of the Sun than are sunspots. These results, although only suggestive, could nevertheless provide useful evidence on the nature and distribution of the effect(s) responsible for this variation and perhaps for the overall deficiency of solar neutrinos.

Discussion

What sort of phenomenon could result in the inferred anticorrelations? As $v \approx \sqrt{T}$, a change in core temperature could yield a first-order change in both the sound-wave travel time, and the neutrino flux. The sign of the frequency shift would then imply, however, that the weighted solar interior temperature is a minimum at the solar minimum. Nevertheless, as the travel time and the p-mode eigenfunctions are strongly weighted outside the core, the temperature probed is weighted more strongly near the surface. Hence a reduced temperature gradient in the solar interior might both reduce the high-energy solar neutrino flux and increase the temperature outside the core. How the necessary solar variations on these timescales could be coupled to the core is not obvious. Alternatively, low-frequency solar oscillations might be explained by the existence of an extremely large (of the order of 10^8 G) non-axisymmetric (predominantly toroidal) magnetic field deep in the Sun, which could provide a restoring force for oscillations¹⁰. This might affect the production of solar neutrinos either by directly affecting core dynamics, or by a mechanism similar to that proposed for surface magnetic fields^{3,4}, but requiring a neutrino magnetic moment that could be three to four orders of magnitude smaller.

Speculation aside, any established correlation between neutrinos and oscillation shifts would imply that a solar variation that is strongly weighted near the surface is coupled to either deep-seated neutrino production, or to neutrino transmission. The features suggested here, such as a correlation with mode number, could provide a means of examining both the angular and the radial profile of the phenomena. If neutrino modulation and a correlation with acoustic oscillations are real, the results of forthcoming gallium solar neutrino experiments will be useful. An unsuppressed production rate of germanium would suggest temperature changes which would predominantly affect high-energy neutrinos whereas a suppressed rate would provide evidence in favour of a neutrino-based explanation. (Preliminary reports from the SAGE experiment are consistent with either possibility.) In any case, the suggestive anticorrelation and the similarities in periodograms displayed here between the ^{37}Cl neutrino event rate and solar acoustic oscillations provide new handles to help understand the nature of the connection between the solar cycle, neutrino production and the solar interior. Further observations, both of p-mode shifts¹¹ and solar neutrinos, are required before the detailed nature of the correlation suggested here, and its implications, can be confirmed and, if real, fully explored. $\square$

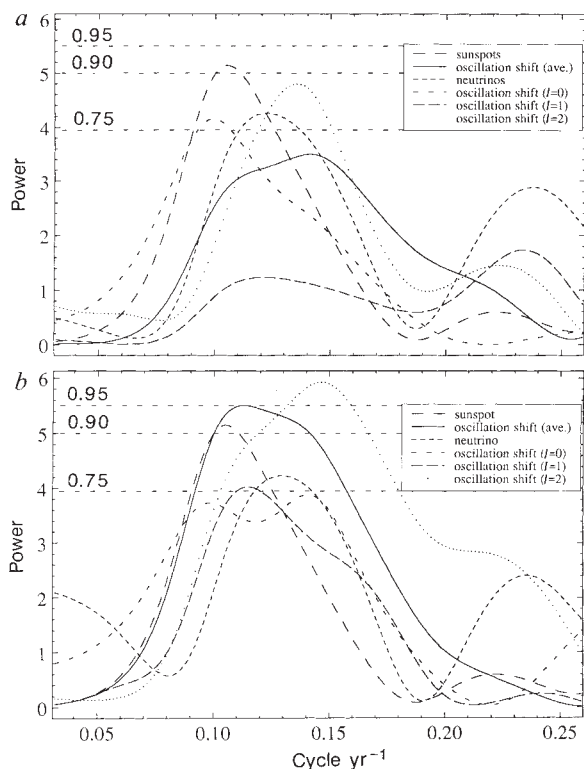


FIG. 4 a, Unweighted and b, weighted periodograms as in Fig. 3, but with the region of the maximum peaks expanded and also showing the periodograms for the individual oscillation mode shifts.

Received 11 July; accepted 11 October 1990.

1. Bahcall, J. N. & Press, W. H. *Solar-Cycle Modulation of Event Rates in the Chlorine Solar Neutrino Experiment*, SNS preprint AST 90/11 (Institute for Advanced Studies, Princeton, 1990).
2. Bieber, J. W., Seckel, D., Stanev, T. & Steigman, G. *Sunspots and Solar Neutrinos* preprint BA-90-36 (Bartol Research Foundation, Newark, 1990).
3. Voloshin, M. B. & Vysotskii, M. I. *Soviet J. nucl. Phys.* **44**, 544–550 (1986).
4. Okun, L. B. *Soviet J. nucl. Phys.* **44**, 546–549 (1986).
5. Elsworth, Y., Howe, R., Isaak, G. R., McLeod, C. P. & New, R. *Nature* **345**, 322–324 (1990).
6. Bahcall, J. *Neutrino Astrophysics* (Cambridge Univ. Press, 1989).
7. Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences* (McGraw Hill, New

York, 1969).

8. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. *Numerical Recipes* (Cambridge Univ. Press, New York, 1986).
9. Press, W. H. & Rybicki, G. *Astrophys. J.* **328**, 277–280 (1989).
10. Dicke, R. H. *Solar Phys.* **78**, 3–16 (1982).
11. Libbrecht, K. G. & Woodard, M. F. *Nature* **345**, 779–782 (1990).

ACKNOWLEDGEMENTS. I thank W. Press for discussions and J. Bahcall for an early preprint. I also thank P. Demarque for comments. This work was supported in part by an NSF Presidential Young Investigator Award and by the US Department of Energy.

Variation of the solar neutrino flux with the Sun's activity

J. W. Bieber*, D. Seckel*, Todor Stanev* & G. Steigman†

* Bartol Research Institute, University of Delaware, Newark, Delaware 19716, USA

† Department of Physics, Ohio State University, Columbus, Ohio 43210, USA

The detection rate of ^{37}Ar at the Homestake solar neutrino experiment is strongly correlated with the monthly sunspot number. The strength of the correlation rises when a semiannual variation is included. These results support the hypothesis that precession of a neutrino magnetic moment in the solar magnetic field is responsible for the variation of the Homestake data.

FOR twenty years the Homestake Neutrino Detector has been monitoring the flux of neutrinos produced in the core of the Sun^{1,2}. The experiment is important because neutrinos are the only known particles to reach Earth directly from the solar core. Thus they provide an observational basis for testing the complex theoretical framework describing the thermonuclear interactions that power the Sun.

Although it is generally accepted that the experiment has succeeded in detecting solar neutrinos, the results continue to be a puzzle in two ways³. First, the inferred neutrino flux is a factor of two to three smaller than the predicted flux. Various explanations for this discrepancy have been put forward, but a definitive resolution has not emerged. The second puzzle is a controversial and unexpected variation of the flux with solar activity. Here we re-examine this effect in light of the latest Homestake data.

The variation is illustrated in Fig. 1, which shows yearly averages of the ^{37}Ar production rate^{1,2}, the sunspot number⁴ and the intensity of $>0.4\text{-GeV}$ cosmic rays registered by the McMurdo (Antarctica) neutron monitor⁵. In discussing the Homestake experiment we will refer to the ^{37}Ar production rate. The ^{37}Ar is produced through the reaction between an electron neutrino and ^{37}Cl , $\nu_e + ^{37}\text{Cl} \rightarrow e + ^{37}\text{Ar}$. The argon is extracted several times a year from a tank containing 133 tons of ^{37}Cl (in the form of C_2Cl_4) and counted by observing the decay of ^{37}Ar , $^{37}\text{Ar} \rightarrow ^{37}\text{Cl} + e^+ + \nu_e$, which has a half-life of 35 days. Traditionally, the solar neutrino flux has been compared to the sunspot number. We also include a comparison with neutron count rates to test the conjecture that some unknown background process associated with cosmic rays (which are known to be modulated by solar activity) might produce a variation in the ^{37}Ar production rate.

The correlation with sunspots

To gain a preliminary quantitative impression of the relative degree of correlation, we calculated unweighted (ignoring error bars) correlation coefficients based on the yearly averages. The

correlation of the ^{37}Ar rate with sunspot number is -0.55 . The confidence level, CL, for this correlation is 0.987, where by confidence level we mean the probability with which the null hypothesis may be rejected. In this case, $1 - \text{CL}$ is the probability that random data would yield a correlation coefficient as large in magnitude as the one calculated. For the neutron count rate, the correlation with ^{37}Ar production is 0.46 (CL 0.959).

The correlation of solar neutrinos with sunspot number is suggestive but not compelling. The correlation with cosmic rays is weaker and could easily arise from the strong association (correlation coefficient -0.78 , CL 0.99996) between sunspots and neutron counts.

A better impression of the association between sunspots and ^{37}Ar rates is given in Fig. 2a, where the ^{37}Ar production rate is displayed for each experimental run as a function of the sunspot number, taken from the month of the mean exposure date for ^{37}Ar production. The solid circles in Fig. 2a show the data grouped into four bins determined by taking the minimum and maximum value of the sunspot number for the 84 individual samples (runs 18 to 108) and forming four equally spaced bins between these limits. The mean ^{37}Ar rate and the error of the mean for each bin were calculated using the larger of the upper and lower error bars quoted in refs 1 and 2. We use this choice of error bars throughout this paper, unless otherwise noted. Figure 2b illustrates a similar reduction for ^{37}Ar production and the neutron-monitor data.

Figure 2a,b reveals, in more dramatic fashion, the same relationships suggested by the correlation coefficients. We have tried a number of different binning schemes, and the correlation evident in Fig. 2 is visible in all cases. Because the number of bins chosen and the method of allocating individual runs into bins is arbitrary, the statistical significance of such procedures is hard to judge. We therefore decided to base our quantitative analysis on the original unbinned ^{37}Ar production rates compared directly with the monthly sunspot number. We measured the significance of the comparison using the F test, as described in ref. 6. This test is appropriate for determining the significance of an improvement in χ^2 when including another parameter in a fit. In our case, we first calculate the mean ^{37}Ar production rate and then test for the significance of including a term describing the variation with sunspot number.

The mean ^{37}Ar production rate is $0.388 \pm 0.031 \text{ atom d}^{-1}$ with $\chi^2 = 90.76$ for the scatter of the data around the mean. We allowed for the possibility that the ^{37}Ar production depends on sunspot number by determining the parameters a_s and b_s in the equation $y = a_s + b_s s$, where y is the ^{37}Ar production rate and s is the sunspot number for the month of mean exposure. The best fit was obtained for $a_s = 0.553$ and $b_s = -0.00185$. The χ^2

for the linear fit is 78.92. The significance of the improvement is indicated by the F statistic, which is the ratio of the improvement in χ^2 (11.84) to the reduced χ^2 of the linear fit (0.9624, which reflects 82 degrees of freedom for the linear fit). For this case the appropriate F statistic is thus 12.3, which indicates, with a formal or 'gaussian' CL 0.9992, that the additional parameter (the slope b_s in this case) is non-zero.

We believe, however, that the CL of 0.9992 derived under the assumption that the data obey gaussian (normal) statistics may overestimate the true confidence level owing to the small number of ^{37}Ar counts in the individual runs—in some cases the most probable physically allowed number of ^{37}Ar atoms produced is zero. These are conditions where gaussian statistical techniques are known to fail. We therefore adopted a procedure for extracting confidence levels empirically, by comparing the value for a statistic derived from the real data to the value for the same statistic derived from a large number of randomly reordered copies of the data⁷. Thus, we perform 10^5 reorderings of the data and find that 394 yield F values larger than that of the original data, corresponding to an empirical CL of 0.996. The empirical $(1 - \text{CL})$ is a factor of five weaker than the gaussian $(1 - \text{CL})$. We quote only empirical confidence levels below. Also, we do not quote error estimates for the parameters a_s and b_s because we have not evaluated them empirically.

It is possible that the variation is not due to a variation in the solar neutrino flux, but to a variation in the background. The primary concern is the possibility that ordinary cosmic rays, by some unknown process, contaminate the ^{37}Ar data and lead to a correlation with solar activity through the modulation of galactic cosmic rays. Such a correlation is not to be expected because the total cosmic-ray flux produces only $0.0003\ ^{37}\text{Ar}$ count d^{-1} by neutrinos generated in atmospheric interactions⁸. Similarly, although very-high-energy (TeV) muons of atmospheric origin penetrate to the neutrino detector and produce a background measured to be $0.08\ ^{37}\text{Ar}$ count d^{-1} (ref. 9), these muons have their origin in ~ 10 -TeV cosmic rays, which are not strongly modulated by solar activity. Nonetheless, perhaps some other vector could mediate between the cosmic rays and the detector. If this were the case then one would expect the neutron-monitor data (Fig. 2b), which directly tracks the cosmic-ray flux, to correlate more strongly with the ^{37}Ar data than do the sunspots. In fact, using the F test we found a weaker result, empirical CL 0.87, and thus conclude that there is no evidence to support this notion.

Seasonal variations

Assuming the correlation between sunspots and solar neutrinos is real, the leading candidate for an explanation concerns the possibility that neutrinos possess small magnetic moments^{10–12}. While traversing the magnetic fields in the upper convection zone, electron neutrinos would precess into neutrinos of some other flavour, which would go unobserved by the ^{37}Cl experiment. If this is the case one might expect some more direct

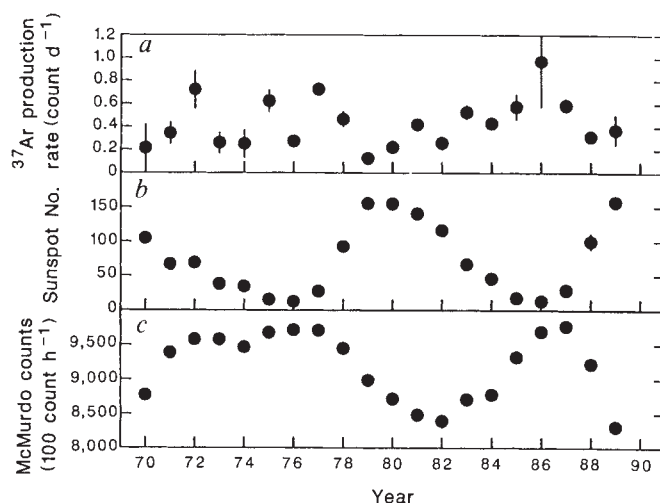


FIG. 1 Yearly means of *a*, the ^{37}Ar production rate measured at Homestake², *b*, the International relative sunspot number⁴ and *c*, the count rate of the McMurdo neutron monitor⁵. The averages of the ^{37}Ar production shown here are based on a sample of 84 runs, with run numbers between 18 and 108, from the Homestake solar neutrino detector. The McMurdo neutron monitor responds principally to galactic cosmic-ray protons with energy > 0.4 GeV. The neutrons detected are secondary particles generated in atmospheric cascades.

indicator of magnetic field strength than sunspot number to yield a stronger correlation. In this spirit, we considered correlations between ^{37}Ar production and the magnitude of the interplanetary magnetic field strength¹³ at 1 AU. There was no evidence for a significant correlation, empirical CL 0.52, which may simply indicate that the field at 1 AU is not a suitable proxy for magnetic fields in the convection zone of the Sun.

In addition to a correlation with sunspots, neutrino precession might also cause a semiannual variation in the rate^{11,12}. The effect has its roots in the fact that the ecliptic is inclined $7^\circ 15'$ to the solar equator. It is generally expected that the toroidal magnetic fields of the Sun are caused by differential rotation. An approximate relationship¹⁴ for small Θ is $\Omega(\Theta) = \Omega(0) - \alpha\Theta^2$, where Ω is the Sun's angular velocity and Θ is heliolatitude. The magnitude of the differential rotation then increases with latitude, $|d\Omega/d\Theta| = 2\alpha|\Theta|$, so the shear amplification of the toroidal field vanishes at the equator, and naively one expects $B_{\text{toroidal}} \propto |\Theta|$ for small Θ . Thus the effects due to neutrino precession should be minimized on 6 June (ascending node) and 8 December (descending node) and the ^{37}Ar production rate should be a maximum on those dates. Similarly, on 7 March and 8 September the ^{37}Ar rate should be a minimum.

To test this idea we studied the relation $y = a_z + b_z|z|$, where

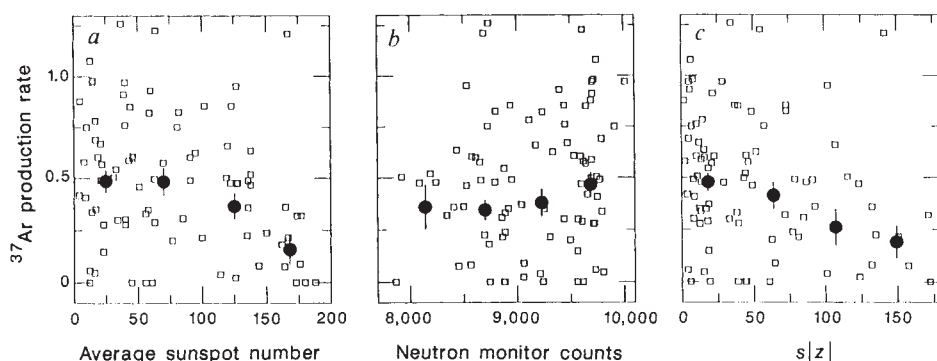


FIG. 2 *a*, ^{37}Ar production rate compared to sunspot number s . The sunspot number associated with an individual ^{37}Ar sample was taken to be the monthly mean for the month in which the mean exposure time of the ^{37}Ar measurement falls. $\square$, individual data points. $\bullet$, mean ^{37}Ar production rate grouped in four equally spaced bins according to sunspot number. *b*, Similar plot for McMurdo count rates. *c*, similar plot for the parameter $B = s|z|$, where z represents Earth's heliolatitude.

$z = \cos(2\pi(\phi - 0.68))$ is proportional to the heliolatitude of the Earth and ϕ is the season of the ^{37}Cl experiment (from 0 to 1; 0.68 corresponds to the date 8 September). This is a very crude treatment of the complicated dynamics of the Sun, but we chose to use z as a simple parameter that still represents the underlying physical ideas. The F test for b_z gave the value 5.8, which has an empirical CL 0.95 that b_z is not zero.

Presumably, the sunspots trace the temporal variations (over 11 years) in the overall strength of the solar magnetic fields, whereas the seasonal variation traces spatial variations in the solar magnetic field. This suggests that the integrated line-of-sight field strength is proportional to both sunspot number and latitude on the Sun. Consequently, we performed the F test for the hypothesis $y = a_B + b_B s|z|$ and found $F = 16.3$ with an empirical CL 0.9993 that b_B is not zero. The best fit is for $a_B = 0.531$ and $b_B = -0.00229$. There is a fivefold improvement in $(1 - \text{CL})$ over that for sunspots alone. The superior correlation obtained by combining the sunspot and seasonal effects can be seen graphically by repeating the binning procedure used in Fig. 2a for sunspots, but now applied to the quantity $B = s|z|$, as illustrated in Fig. 2c.

We looked at two other possible seasonal variations. If the Mikheyev-Smirnov-Wolfenstein (MSW) effect^{3,15} is the explanation for the overall low production rate of ^{37}Ar then there should be, at some level, an annual (not semiannual) variation in the ^{37}Ar production rate due to 'back oscillations' in the Earth¹⁶. In the Northern Hemisphere the production rate should be a maximum at the winter solstice. Coincidentally, owing to the 0.0167 eccentricity of the Earth's orbit, there should be a maximum in the solar neutrino flux at perihelion (1 January), which would reinforce any effect of back oscillations. We looked for these effects by testing the hypothesis $y = a + b_{\text{annual}} \cos(2\pi(\phi - 0.97))$, where 0.97 corresponds to the winter solstice, and found $F = 0.16$, a null result.

Lies, damn lies and statistics

We have concluded that the neutrino flux measured by the Homestake experiment exhibits a strong correlation with sunspots, which is modestly enhanced when seasonal variations are included. This is controversial enough to merit a further examination of the statistical techniques underlying our conclusions. In particular, we wish to provide the skeptical reader a basis for judging how meaningful are our quoted confidence levels.

As stated earlier, we believe that the formal (gaussian) confidence level of the F statistic is too high due to the low number of counts and the inappropriateness of gaussian statistics. We have therefore used a system of empirical confidence levels based on random reorderings of the data. In general, we find that the empirical CL is substantially weaker than the gaussian CL whenever CL is high. This is illustrated in Fig. 3, where we show the distribution of gaussian CL for F calculated from 10^5 reorderings of the ^{37}Ar data with respect to B . The arrow shows the formal gaussian CL for the correct ordering. Clearly, the distribution is significantly wider than it should be if gaussian statistics were applicable. For example, with 10^5 reorderings one would expect $1,000 \pm 31$ to exceed CL 0.99, where in fact we find 2,715.

It is conceivable that the empirical confidence levels are biased because they are derived from reorderings of the real data set, which may have some peculiar features (not least of which is the purported variation with sunspots). To test this possibility we obtained a set of 249 simulated data sets for the Homestake experiment (B. Cleveland, personal communication). These data sets use the same run times, extraction dates, counter backgrounds and analysis techniques as the real data. The number of ^{37}Ar events for each run is determined using a Poisson distribution appropriate for the exposure time and a constant ^{37}Ar production rate at a level equal to the measured mean rate. A histogram of the gaussian CL for the simulations is also shown

in Fig. 3. Again the distribution is broader than gaussian statistics would indicate. There are 8 cases in which CL 0.99, is exceeded where we would expect 2.5 ± 1.6 . Given that the number of trials is a factor of 400 smaller than for the random reorderings of the real data, this is a comparable excess, which suggests that our empirical confidence levels are unbiased. We also note that none of the simulations yielded a correlation with B better than that obtained for the real data, consistent with our empirical confidence level of 0.9993.

Another worry concerns the uncertainties reported by the Homestake collaboration, which are not error bars on the number of counts, but rather maximum likelihood limits on the inferred production rate for the individual runs. Thus, although there is considerable scatter, the size of the error bars is correlated with the magnitude of the data, with smaller ^{37}Ar production rates receiving higher weight in a gaussian analysis. The problem is illustrated by the mean ^{37}Ar production rate. Calculated using gaussian statistics, our result of 0.388 ± 0.031 atom d^{-1} is significantly smaller than the Homestake collaboration's maximum likelihood estimate of 0.52 ± 0.05 atom d^{-1} . This discrepancy suggests that the correlation of small errors with small ^{37}Ar production rates might be responsible for the broad F distribution. To test this, we 'de-correlated' the data and error bars by reordering them independently. The distribution of F is, however, similar to that in Fig. 3. Assigning all error bars a constant value did make a difference. The distribution of empirical CLs narrowed to be consistent with the gaussian CL. Further, the CL for the real data became weaker, qualitatively consistent with the unweighted correlation of yearly

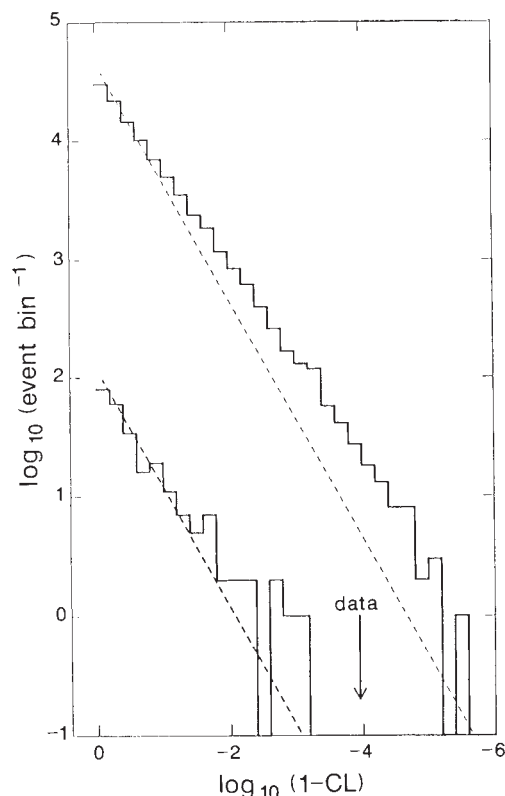


FIG. 3 Results of two numerical experiments to test the applicability of gaussian statistics for determining the confidence level (CL) of a correlation between ^{37}Ar production and the parameter $B = s|z|$. In the upper histogram, we show gaussian confidence levels determined from the F statistic for 10^5 randomly reordered copies of the real data set. The distribution is wider than expected if gaussian statistics were valid, as shown by the dashed line. To calculate an empirical CL, we integrated to the left of the gaussian CL for the correct ordering of the real data (arrow) and normalized by the complete integral. For the lower histogram we used 249 simulated data sets for which a constant solar neutrino flux was assumed.

averages. Neglecting the error bars, however, ignores information concerning counter backgrounds, exposure times and so on and, we believe, is unjustified.

Together, these tests lead us to believe that the broadening of the F distribution occurs because gaussian statistics are not strictly applicable, and not because of the correlation of error bars with production rates or some peculiar feature of the real data. If so, then a complete treatment of the confidence levels from first principles would require more detailed access to the data than was available for this analysis.

Discussion

Before this year, the most recent analysis of time variation in the ^{37}Ar production rate⁷ included data through 1985. Using a technique similar to our empirical confidence levels, a CL of 0.95 was reported.

Recently, the question of time variation, using data through run 105, has been discussed in two other papers^{17,18}. The statistical techniques are different from ours, so a comparison of confidence levels, although interesting, is hazardous. Bahcall and Press¹⁷ evaluated rank-order statistics, dispensing with the question of error bars entirely, and found CLs, by two different methods, of 0.987 and 0.991 that the ^{37}Ar production rate is correlated with sunspot number. Filippone and Vogel¹⁸, on the other hand, address the inadequacy of gaussian statistics by regenerating the raw number of ^{37}Ar counts under the assumption that the error bars are proportional to the square root of the number of events. Using a reordering technique they found an empirical CL of 0.993 that a correlation with sunspots could not be produced by random fluctuations of a constant ^{37}Ar production. It is also possible to extract an F statistic from their results which should obey gaussian statistics. We infer a gaussian CL 0.991 that ^{37}Ar production is correlated with sunspots.

Our results, in conjunction with those of refs 17 and 18, show that the inclusion of ~ 15 more runs of data has improved the correlation with sunspots by a factor of 5–10. This is illustrated in Fig. 4, where we plot the empirical confidence levels based on 10^5 reorderings for subsets of the data consisting of just the first n runs. Up to and including run number 59 there is no apparent correlation. The next four runs, as pointed out in

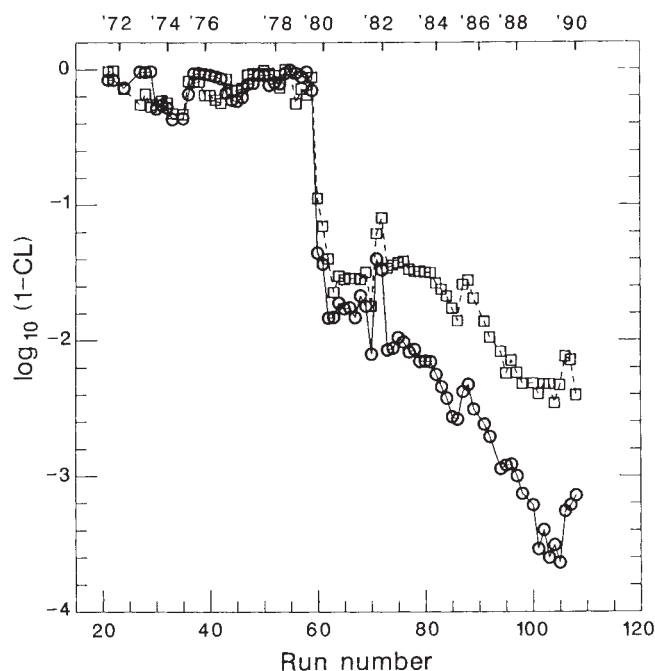


FIG. 4 Empirical confidence level for the correlation of Homestake data with sunspot number ($\square$) and $B=s|z|$ ($\circ$), plotted as a function of the number of runs included in the analysis. The beginning of calendar years is indicated on the upper scale.

Bahcall *et al.*⁷, account for much of the correlation through 1985, and one might ask if those runs were atypical in some way. The data in the last four years improves the correlation significantly, however, whereas one would expect a constant neutrino rate to gradually reduce the effects of an accidental excursion. Also, Bahcall and Press¹⁷ noted that the last two-thirds of the data correlates much better with sunspots than does the first third. Figure 4 suggests that it is really the data from run number 60 onward that correlates well. Remarkably, the time of run 60 coincides closely with the time of polarity reversal in the 22-year solar magnetic cycle, but we have no idea whether or not this has physical significance.

We have concentrated on the temporal variation in ^{37}Ar production; however, such effects cannot be considered in isolation. Thus, although we have emphasized the neutrino-magnetic-moment hypothesis, it must be admitted that there are problems with this interpretation. First, the magnetic-moment hypothesis cannot by itself explain the overall low rate of production of ^{37}Ar . In the simplest model of spin precession the maximum suppression would be a factor of one-half when the neutrino beam is totally depolarized. The latest predictions for ^{37}Ar production are 7.9 ± 2.6 (3σ) solar neutrino units (SNU)¹⁹ and 5.8 ± 1.2 (1σ) SNU²⁰, whereas the mean ^{37}Ar production rate corresponds to 2.33 ± 0.25 SNU¹, less than half the prediction. Alternatively, using a_B for the maximum production rate, and including a background subtraction, yields 2.36 SNU. Even allowing this value to increase by 30% for the inadequacy of gaussian statistics gives a ^{37}Ar production rate that is $\geq 2\sigma$ smaller than the lowest prediction. Second, for these simple models the precession is independent of neutrino energy, and the sunspot and seasonal variations should show up in any detector sensitive primarily to electron neutrinos. The Kamiokande detector^{21,22} has also been monitoring the solar neutrino flux since 1987 and sees no obvious sign of time variation, although we feel that three years is too short a time for a definitive conclusion. Even so, there is not necessarily a contradiction²³ because the Homestake experiment detects only electron neutrinos absorbed on ^{37}Cl , whereas Kamiokande detects scattered electrons, which may be produced by neutrinos of any flavour through weak interactions, or through electromagnetic interactions if neutrinos do indeed possess a magnetic dipole moment. A third problem is that sufficiently large neutrino magnetic moments would cause an apparent inconsistency with the theory of red giant stars²⁴.

Considering all the evidence, it seems that a full solution of the solar neutrino puzzle requires something in addition to the hypothesis of neutrino magnetic moments. Whether this will manifest itself as a change to the standard solar model, or perhaps a combination of the magnetic moment hypothesis with the MSW effect²⁵, is something that can only be determined by performing other experiments with different sensitivity to the solar neutrino flux³.

Summary

We have found strong indications that the ^{37}Ar production rate varies in concert with sunspots and seasonal factors. The confidence level for a correlation with sunspots is 0.996, including data through run 108. This compares with two other recent analyses which both yield 0.991 for data through run 105, and a previous analysis including data from 61 runs which yields CL 0.95. Although the techniques in these analyses differ, the correlation between ^{37}Ar production and sunspots has become stronger over the last five years.

We also find a weaker semiannual variation tied to the absolute value of Earth's heliolatitude. A first attempt to combine the heliolatitude dependence with the sunspot correlation yields a confidence level of 0.9993, suggesting that the ^{37}Ar production rate at the Homestake neutrino detector is correlated with subsurface solar magnetic fields in accordance with the predictions of models involving neutrino magnetic moments. $\square$

Received 6 June; accepted 12 October 1990.

1. Davis, R. Jr et al. *Proc. 21st Int. Conf. Cosmic Rays* Vol. 12 (ed. Protheroe, R. J.) 143–151 (Physics Publications, University of Adelaide, 1990).
2. Rowley, J. K., Cleveland, B. T. & Davis, R. Jr in *Proc. Conf. Solar Neutrinos and Neutrino Astronomy* Vol. 126 (eds Cherry, M. L., Lande, K. & Fowler, W. A.) 1–21 (Am. Inst. Phys., 1985).
3. Bahcall, J. N., Davis, R. Jr & Wolfenstein, L. *Nature* **334**, 487–493 (1988).
4. Coffey, H. E. (ed.) *Solar-Geophysical Data, Rep. No. 540-Part 1* (National Geophysical Data Center, Boulder, 1989).
5. *Cosmic Ray Nucleonic Intensity Data* (Bartol Research Institute, Newark, Delaware, 1970–1989).
6. Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences* (McGraw Hill, New York, 1969).
7. Bahcall, J. N., Field, G. B. & Press, W. H. *Astrophys. J.* **320**, L69–L73 (1987).
8. Gaisser, T. K. & Stanev, T. in *Proc. Conf. Solar Neutrinos and Neutrino Astronomy* Vol. 126 (eds Cherry, M. L., Lande, K. & Fowler, W. A.) 277–281 (Am. Inst. Phys., 1985).
9. Fireman, E. L., Cleveland, B. T., Davis, R. Jr & Rowley, J. K. in *Proc. Conf. Solar Neutrinos and Neutrino Astronomy* Vol. 126 (eds Cherry, M. L., Lande, K. & Fowler, W. A.) 22–31 (Am. Inst. Phys., 1985).
10. Cisneros, A. *Astrophys. Space Sci.* **10**, 87–91 (1971).
11. Voloshin, M. B., Vysotskii, M. I. & Okun, L. B. *Soviet J. nucl. Phys.* **44**, 440–441 (1986).
12. Voloshin, M. B. & Vysotskii, M. I. *Soviet J. nucl. Phys.* **44**, 544–545 (1986).
13. King, J. H. *Interplanetary Medium Data Book Suppl.* 4 (1985–1988) (National Space Science Data Center, Greenbelt, 1989).
14. Howard, R. A. *Rev. Astr. Astrophys.* **22**, 131–155 (1984).
15. Mikheyev, S. P. & Smirnov, S. *Nuovo Cim.* **C9**, 17–26 (1986).
16. Cribier, M., Hampel, W., Rich, J. & Vignaud, D. *Phys. Lett. B* **182**, 89–94 (1986).
17. Bahcall, J. N. & Press, W. H. *Astrophys. J.* (submitted).
18. Filippone, B. & Vogel, P. *Phys. Lett. B* **246**, 546–550 (1990).
19. Bahcall, J. N. & Ulrich, R. K. *Rev. mod. Phys.* **60**, 297–372 (1988).
20. Turck-Chieze, S., Cahen, S., Casse M. & Doom, C. *Astrophys. J.* **335**, 415–424 (1988).
21. Hirata, K. S. et al. *Phys. Rev. Lett.* **63**, 16–19 (1989).
22. Hirata, K. S. et al. *Phys. Rev. Lett.* **65**, 1297–1300 (1990).
23. Suzuki, A. KEK Preprint 90-31 (1990).
24. Raffelt, G. Max Planck Institute preprint MPI-PAE/PTH 6/90 (1990).
25. Lim, C.-S. & Marciano, W. J. *Phys. Rev. D* **37**, 1368–1373 (1988).

ACKNOWLEDGEMENTS. We thank K. Lande and R. Davis Jr for discussions and for unpublished data, and B. Cleveland for simulation data. We thank B. Filippone for pointing out the problem with the mean ^{37}Ar production rate. This work is supported by the NSF and the Department of Energy.

Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains

Jiahui Wang*, Youwei Yan*†, Thomas P. J. Garrett*†, Jinhuan Liu*, David W. Rodgers*, Robert L. Garlick‡, George E. Tarr§, Yasmin Husain§, Ellis L. Reinherz§|| & Stephen C. Harrison*†

* Harvard University, Department of Biochemistry and Molecular Biology, and † Howard Hughes Medical Institute, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

‡ Upjohn Company, Kalamazoo, Minnesota 49007, USA

§ Laboratory of Immunobiology, Dana Farber Cancer Institute and || Department of Medicine, Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA

The structure of an N-terminal fragment of CD4 has been determined to 2.4 Å resolution. It has two tightly abutting domains connected by a continuous β strand. Both have the immunoglobulin fold, but domain 2 has a truncated β barrel and a non-standard disulphide bond. The binding sites for monoclonal antibodies, class II major histocompatibility complex molecules, and human immunodeficiency virus gp120 can be mapped on the molecular surface.

THE cell-surface glycoprotein CD4 is expressed on most thymocytes and on the subset of peripheral T lymphocytes that includes helper T cells^{1,2} and class II major histocompatibility complex (MHC) specific cytotoxic T cells^{3–5}. Several experiments suggest that CD4 contacts nonpolymorphic regions of class II MHC molecules and that this interaction is important for triggering through the T-cell receptor by class II MHC-presented antigen^{3–8}. The CD4/class II MHC interaction is also critical for a key stage in thymocyte differentiation⁹. CD4 may play a direct part in T-cell activation¹⁰, and a specific tyrosine kinase associated with the cytoplasmic domain of CD4 provides a possible route for ligand-dependent signal transduction^{11,12}.

Human CD4 is the receptor for HIV^{13–18}. Binding of the viral glycoprotein, gp120, to CD4 is the initial step in viral entry, leading to fusion of viral and cell membranes^{19,20}. Blocking viral attachment by inhibiting gp120–CD4 binding is a potentially valuable antiviral strategy in AIDS therapy.

CD4 (relative molecular mass 55,000) consists of an extracellular portion (residues 1–371), a transmembrane segment (372–395), and a cytoplasmic tail (396–433) (refs 21, 22). The N-terminal 100 residues have clear sequence similarity to immunoglobulin variable domains (particularly V κ) (ref. 22), and mapping of sites for anti-CD4 monoclonal antibodies indi-

cates that this part indeed folds like an immunoglobulin variable region^{23,24}. Less striking similarities suggest that residues 100–180, 180–290 and 290–370 (roughly) may also be homologous to immunoglobulin domains^{22,25}. The extracellular portion has therefore been described as a concatenation of four immunoglobulin-like units. We use the domain terminology here, because our structure bears out predictions about domains 1 and 2, but evidence for distinct third and fourth domains is only indirect. Recombinant products corresponding to the intact extracellular region ('domains 1–4') are monomeric⁴¹, and the CD4 glycoprotein is believed to be monomeric on the cell surface as well²⁶. The two N-glycosylation sites on human CD4 are in domains 3 and 4²¹. HIV gp120 binds to various soluble CD4 derivatives^{27–33}. The known critical contacts are confined to domain 1 (refs 23, 33–37). Contacts to class II MHC molecules probably extend over domains 1 and 2 (refs 38, 39).

Several groups have obtained crystals of recombinant soluble CD4 derivatives^{40–42}. Those from four-domain species have large unit cells, and they diffract relatively poorly. By contrast, crystals of N-terminal, two-domain products diffract to high resolution. We report here on a structure obtained from two forms of the two-domain species. Independent results on similar crystals are reported in the accompanying paper by Ryu *et al.*⁴³.

Structure determination

We studied two forms of truncated CD4, referred to here as CD4(1–182) and CD4(1–183). CD4(1–182) was produced in secreted form by transfected Chinese hamster ovary (CHO) cells; it contains the first 182 residues of the intact molecule. CD4(1–183) was produced as an inclusion-body protein in *Escherichia coli* using the vector pTrp2-CD4/183; it contains an N-terminal methionine and 183 residues of authentic CD4. It was refolded as described in ref. 44. Both products were purified by immunoaffinity chromatography and gel filtration^{29,44}.

Three crystal forms were identified, all in space group C2, and called A, B and C. Both products seem to give all three

forms. We focused on the C form of CD4 (1–182), which diffracts to spacings of at least 2.2 Å. This report is based primarily on a determination of the 2.4 Å resolution structure from those crystals, using multiple isomorphous replacement (MIR), followed by iterative cycles of refinement and model building. We also solved the structure of the A form of CD4 (1–183) to 3 Å resolution by molecular replacement, and we used averaging of density in the two unit cells to resolve ambiguities at late stages in the refinement of the C-form structure.

Aspects of the C-form structure determination are summarized in Table 1a. We identified three heavy-atom derivatives (K_3OsCl_6 , K_2PtBr_4 and UO_2SO_4) by soaking under varied conditions of pH and ionic strength, to minimize non-isomorphism. The Os and Pt derivatives were solved by examination of isomorphous and anomalous-difference Patterson functions; the uranyl derivative, by a difference Fourier synthesis. Introduction of the heavy atoms greatly reduced the attainable resolution, as shown in Table 1a. Moreover, non-isomorphism, especially of the uranyl derivative, seemed to limit the accuracy of the phase determinations, and the MIR map was of marginal quality (Fig. 1). Nonetheless, clear density for two lobes, each with β sheets, could be seen. Solvent flattening⁴⁵, using an envelope determined by visual inspection and using the Bricogne program suite for density modification⁴⁶, improved the clarity of the map. The expectation that one part of the molecule should resemble an immunoglobulin variable domain enabled us to fit most of domain 1, including many side chains, using the program FRODO⁴⁷, and we could then build segments of a polyalanine model into parts of domain 2, for which the density was weaker.

About 60% of the polypeptide chain, with about half the atoms, was fitted at this stage and preliminarily refined at 3.5 Å resolution using simulated annealing in the program XPLOR^{48,49}. MIR and partial structure phases were combined⁴⁶, Fourier coefficients were computed with modified Sim weights⁵⁰, and solvent regions in the phase-combined map were flattened, as described above. Three rounds of model building, refinement, and phase combination, increasing the resolution at each stage, led to an essentially complete structure at 2.8 Å resolution. This model revealed the immunoglobulin topology of domain 2. Four rounds of refinement and rebuilding, using omit-refine and difference maps to adjust the model and to add some remaining residues, extended the resolution to 2.6 Å. Individual thermal parameters were then introduced, and the resolution was extended to 2.4 Å in several more rounds of refinement and local rebuilding, with the aid of maps produced by density-averaging between A-form and C-form structures, as described below. This averaging not only resolved ambiguities in certain surface loops, but also reduced bias in those regions for which the MIR maps had not provided a clear guide for initial model building. Refinement parameters appear in Table 1c. The *R* factor (using $F > \sigma$) is 22% for data between 6 and 2.4 Å. Ordered solvent has not so far been included.

The r.m.s. error in atomic positions of 0.2 Å was estimated by comparing coordinates before and after a last round of molecular dynamics at 300 K. The errors are not evenly distributed throughout the structure, however. The least well determined residues have main chain atoms with large thermal parameters ($>50 \text{ Å}^2$); these are residues 20–23, 103–107, 124–126,

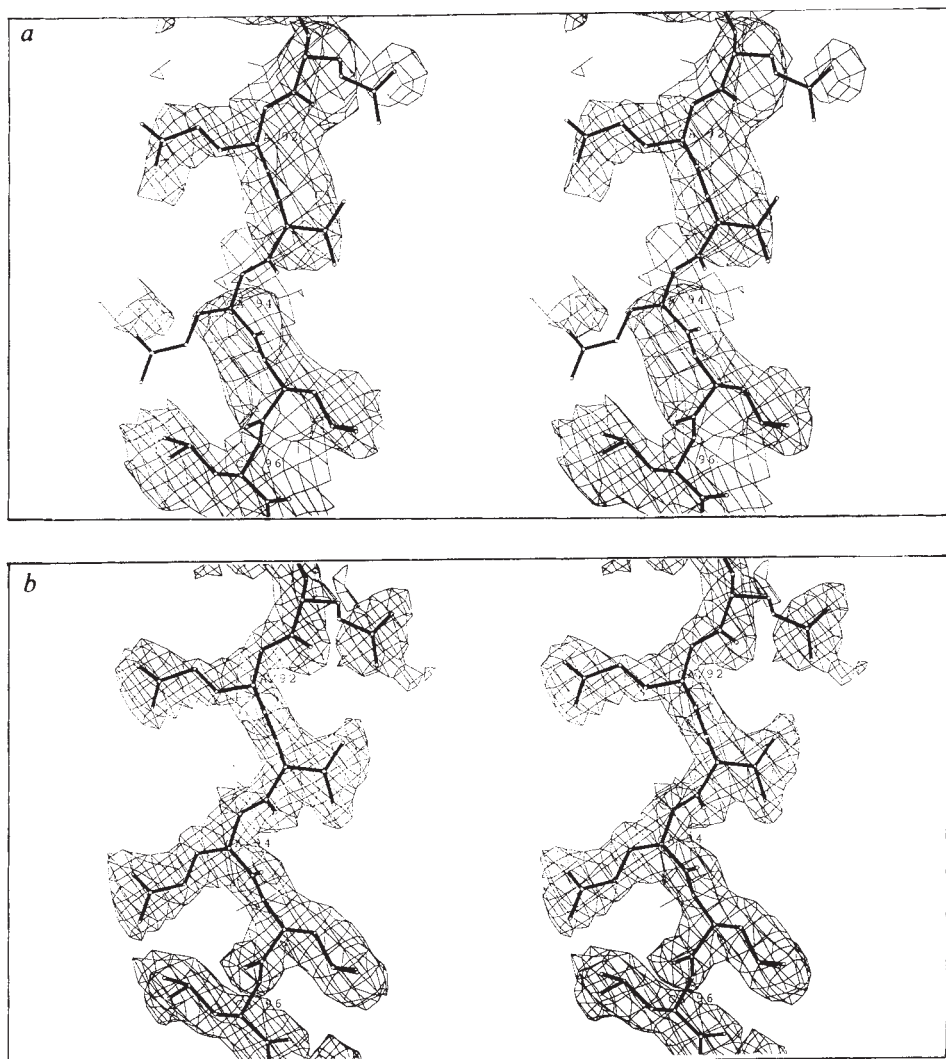


FIG. 1 Portion of the C-form electron density map, showing density for residues 92–96 at two stages of the structure determination. *a*, MIR map, 3.5 Å resolution. *b*, $(2F_o - F_c)$ map after the last round of refinement reported here, 2.4 Å resolution.

131–138 and 151–154. Little or no electron density is present for residues beyond 176, and the side-chain positions for residues 1, 24, 43, 129, 139 and 150 are also uncertain. As indicated by early electron density maps, the average thermal parameters for domains 1 and 2 differ substantially ($20 \text{ \AA}^2$ and $43 \text{ \AA}^2$, respectively), but this difference seems less dramatic in the A form.

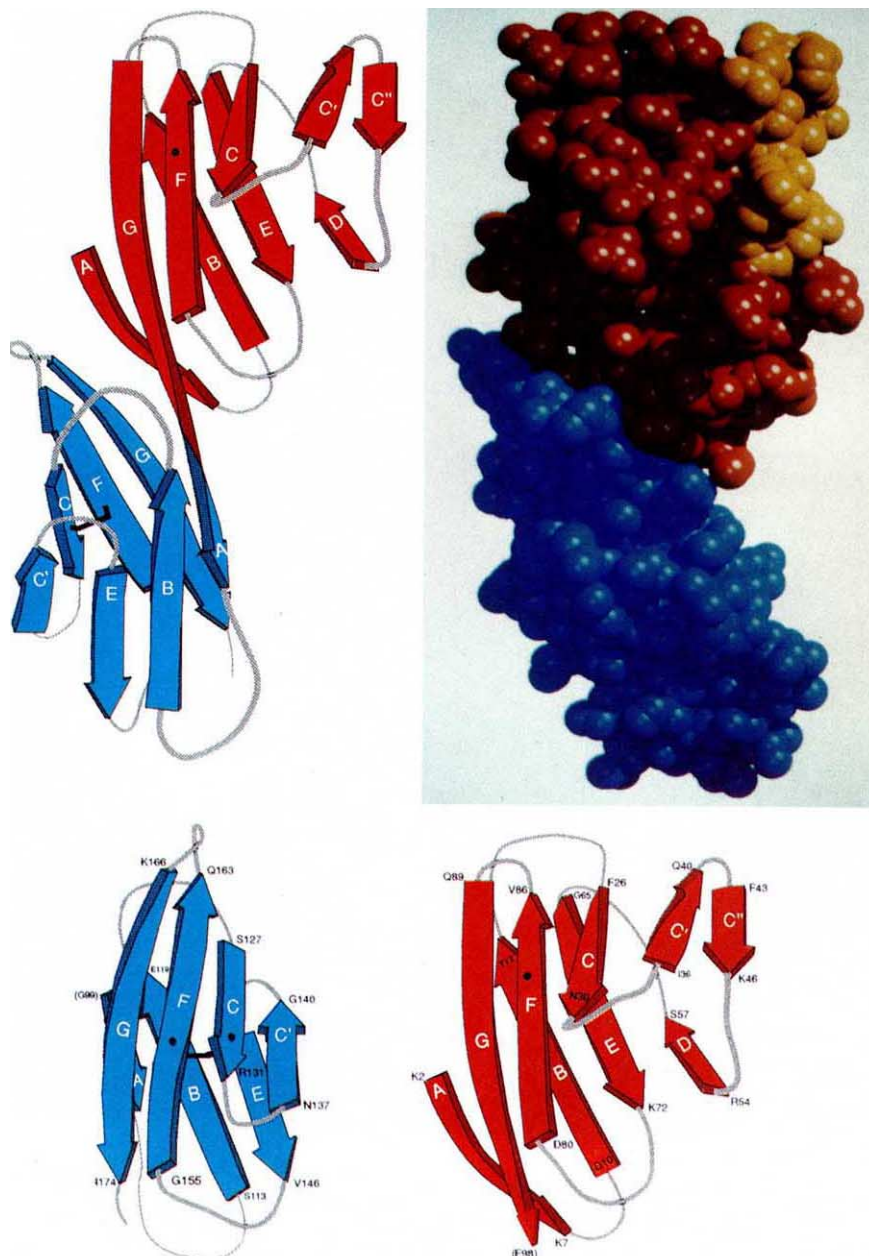
The structure of CD4(1–183) in the A form was solved by molecular replacement using data to $3.3 \text{ \AA}$ resolution (Table 1b) and a model from CD4(1–182) in the C form at an intermediate refinement stage. The fast rotation function⁵¹ and a Patterson-space rotation search⁵² gave identical solutions. A translation search, followed by rigid-body refinement with either CORELS⁵³ or XPLOR⁴⁹, gave $R = 39\%$ to $3.3 \text{ \AA}$. Computations in which the model was split into two domains gave no relative reorientation, showing absence of hinge-bending at the junction. One round of simulated annealing refinement with XPLOR reduced the R factor to 25% for data $7\text{--}3 \text{ \AA}$. Density averaging between the two space groups, using data to $3 \text{ \AA}$ resolution and carried out as described for HLA⁵⁴, gave significant phase improvement, and maps calculated with those phases substantially accelerated

completion of the C-form refinement. The C-form model at a late stage was then used as a starting point for a round of simulated annealing refinement in the A form (to $3 \text{ \AA}$), resulting in an R factor of 20% for a partially refined structure with good geometry (see Table 1c). Successful refinement and averaging show that CD4(1–183) is essentially identical in structure to CD4(1–182).

Molecular structure

The polypeptide chain of CD4(1–182) folds into two abutting domains (Fig. 2). Each domain is an antiparallel β barrel with the connectivity characteristic of an immunoglobulin fold. The two domains pack so closely, one against the other, that there is a large common hydrophobic interface. Moreover, the strand from residue 89–104 passes continuously from one domain to the next, so that residue 98 forms a hydrogen bond in a sheet of domain 1 and residue 99 forms a hydrogen bond in a sheet of domain 2. The overall dimensions are roughly $65 \times 35 \times 25 \text{ \AA}$, giving the molecule a rod-like shape. The N terminus is near one end and the C terminus near the other. Evidence from electron microscopy⁴² and hydrodynamic measurements⁴¹ and

FIG. 2 *a*, Backbone representation of CD4(1–182). Domain 1 is in red, domain 2 in blue; β strands are indicated by letters, separately in each domain. Strand A of domain 2 is continuous with strand G of domain 1. Note that domains 1 and 2 are related by a rotation of $\sim 160^\circ$ and a translation along the axis of the molecule. Disulphide bonds are shown as solid lines; only the trace is visible of the disulphide bond between strands B and F in domain 1. *b*, Solid representation of CD4(1–182), in an orientation similar to *a*. The C' ridge of domain 1, implicated in the binding of HIV gp120, is highlighted. *c*, Representations of domains 1 and 2 oriented to show the similarity of their folded structures. First and last residues in each strand are indicated by single-letter code and sequence numbers.



inferences from crystal-packing analyses of CD4(1-375) (refs 41, 42) suggest that the full extracellular segment is a rod-like molecule about 125 Å long and 25-30 Å wide. The shape of CD4(1-182) is clearly appropriate for half of such a rod, with the C terminus of its polypeptide chain at the membrane-proximal end. The orientation chosen for the illustrations therefore places the membrane below the model. The parts of the molecule at the top of the illustrations can be assumed to be the most accessible on the cell surface.

Domain 1. As predicted from sequence alignments, the structure of the first domain resembles very closely the variable domain of an immunoglobulin light chain. Fig. 3a shows a comparison with the V κ domain of the Bence-Jones protein REI⁵⁵, and Fig. 4 presents the sequence alignment corresponding most closely to this three-dimensional superposition. Immunoglobulin variable domains have two β sheets, with strands conventionally labelled ABCC'C''DEFG^{56,57}. Strands ABED form one sheet; strands GFCC'C'', the other. Strand A actually begins in the

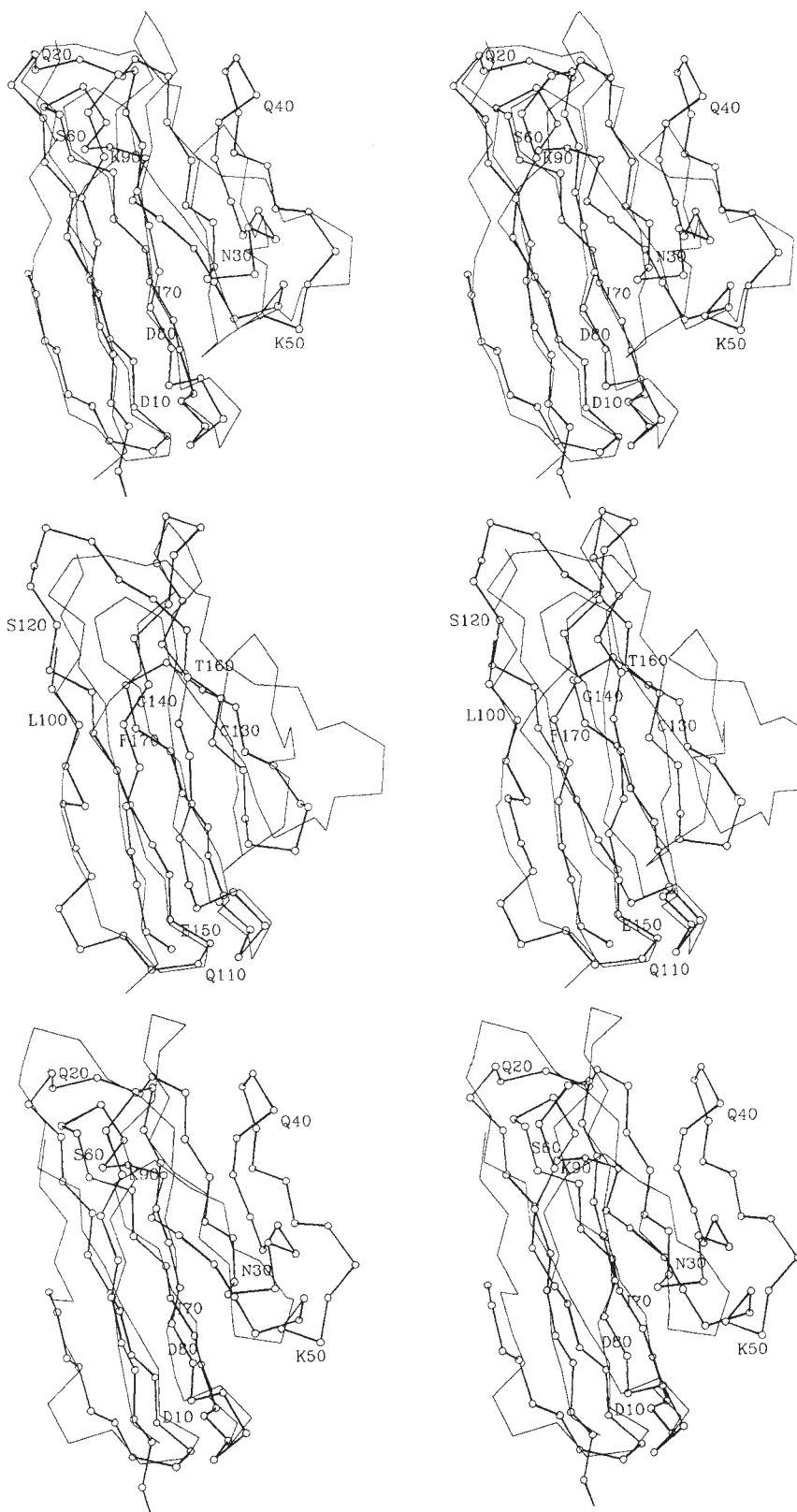


FIG. 3 Stereo drawings of pairwise α -carbon backbone superpositions of CD4 domain 1 (residues 1-98), CD4 domain 2 (residues 99-176), and immunoglobulin V κ domain of Bence-Jones protein REI⁵⁵. *a*, CD4 domain 1 (thick lines and residue numbers) versus REI (thin lines). The drawing is based on fitting the 53 common atoms in the so-called Pin region (ref. 56). The r.m.s. fit is 0.83 Å, comparable to pairwise comparisons among various immunoglobulin domains⁵⁶. *b*, CD4 domain 2 (thick lines and residue numbers) versus REI (thin lines). The fit is based on 42 C α atoms from strands A, B, E, F and G, selected by visual inspection. The r.m.s. agreement is 1.45 Å. *c*, CD4 domain 1 (thick lines and residue numbers) versus domain 2 (thin lines). The fit is based on 38 corresponding C α atoms in strands A, B, C, E, F and G. The r.m.s. agreement is 1.28 Å.

first sheet, alongside B, and passes across to the second, alongside G. The first residue of CD4 corresponds to the residue of REI immediately after this switch, and the two sheets of CD4 domain 1 therefore contain strands BED and AGFCC'C" respectively.

As in all standard immunoglobulin domains^{58,59}, strand B contains the first cysteine (residue 16) of a disulphide, and strand C contains a conserved tryptophan (residue 28) packed against it. The turn from C to C' is more abrupt than in REI and the C'C" corner correspondingly more extended. This extension of the C'C" β ribbon is essentially a β turn. Its high thermal parameters imply some flexibility. It packs against the side chain of Trp 62, which projects from an insertion (relative to REI) forming a helical turn in the DE loop. In heavy-chain variable domains, the size of the framework residue at position 71 in the conventional numbering scheme has been shown to be an important determinant of the conformation of the C'C" loop in CDR2⁶⁰. In CD4, the side chain of Trp 62 occupies the space corresponding to heavy-chain residue 71, and the large tryptophan ring clearly holds the C'C" turn in its outward-projecting position. With Leu 44, in the C" strand, and Phe 67, in the DE loop, Trp 62 forms an extension of the immunoglobulin-like hydrophobic core. Through its effects on the shape of the C'C" side of the domain, this extended hydrophobic core seems to be important for interactions with class II MHC molecules and HIV gp120 (see below).

The first residue of strand D is an arginine, conserved in immunoglobulin variable domains. It forms a salt bridge to a conserved aspartic acid in the EF corner. These residues in the CD4 sequence have been noted previously and cited as evidence for a variable-domain-like folded structure. Another conserved residue, Tyr 82, lies in strand F, with its side chain in the hydrophobic core, forming a hydrogen bond to the main-chain carbonyl of Asp 78. Strand F also contains the second cysteine (residue 84) of the characteristic immunoglobulin disulphide. The FG corner is five residues shorter than in REI. This loop corresponds to the CDR3 region of an immunoglobulin variable domain, where it is especially prominent and contributes critically to antigen binding. Since antigen binding is not a function of CD4, it is not surprising that the polypeptide chain makes an abrupt β turn at residues 87–88, rather than maintaining a large surface projection.

Domain 2. The second domain proceeds directly from the first. There is no equivalent of an elbow or hinge, and the A strand in domain 2 is a direct continuation of the G strand in domain 1. This continuous β -strand connector shows how a stiff, rod-like molecule can be constructed of sequential immunoglobulin-like domains. The structure of domain 2 deviates from a conventional immunoglobulin domain in two important respects: the disulphide links cysteines within one sheet rather than across the barrel, and there is a short connection between the end of strand C and the beginning of strand E. We have called the central part of this connection C', because it participates in four antiparallel β -sheet hydrogen bonds with strand C and therefore corresponds roughly to strand C' of a variable domain rather than to strand D of a constant domain. Indeed, much of the three-dimensional structure of domain 2 superposes well on the first domain and on the V κ domain of REI. The comparisons are shown in Fig. 3b, c, and the sequence alignments derived from these superpositions are given in Fig. 4. Segments of domain 2 that correspond closely to segments of REI are the beginning of the A strand where it runs alongside B, the second part of the AB loop, the B strand and part of the C strand, the entire segment from the beginning of the E strand to the end of the F strand, and the second part of the G strand. The AB and EF loops superpose well on REI, including a conserved glycine in AB and a conserved aspartic acid in EF, both characteristic of immunoglobulin variable domains⁵⁹. One model for domain 2 based on sequence alignments accurately assigned

residues to strands and correctly predicted the placement of the disulphide²⁵.

The first cysteine of the disulphide (residue 130) is in strand C, next to its partner (residue 159) in strand F. It occupies the position taken by the nearly invariant tryptophan in immunoglobulin structures. Disulphides between hydrogen-bonded residues in antiparallel β strands are rare, because the positioning is unfavourable for the required geometry of the S-S bridge⁶¹, but sequence alignments do predict side-by-side disulphide bonds in other members of the immunoglobulin superfamily²⁵. The cysteine residue in strand F corresponds to the 'usual' position for the second member of the disulphide. In effect, domain 2 of CD4 has lost the cysteine in strand B and acquired one in strand C. A different disulphide migration seems to have occurred in the α chain of CD8⁶². Three cysteines are found in its immunoglobulin-like domain, two in the 'usual' locations and a third two residues before the conserved tryptophan in strand C. It has been reported that the disulphide in this domain links this 'extra' cysteine in strand C with cysteine in strand B, leaving the one in strand F unpaired⁶². Thus CD4 and CD8 both seem to have immunoglobulin-like domains with unconventional disulphide linkages. The third domain of CD4, not present in the 1–182 fragment, is believed to have an immunoglobulin-like structure with no disulphide bond²⁵. Precedent for such a structure is found in the two immunoglobulin-like domains of the *E. coli* chaperone protein, pap D⁶³.

The unusual placement of the disulphide bond produces little distortion in domain 2 with respect to REI or domain 1, and despite extensive differences in the identity of particular residues, the overall shape of the hydrophobic core is maintained. The placement of the cysteine in strand C eliminates the normally conserved tryptophan, and a leucine occupies the position of the standard cysteine in strand B. There is a tryptophan instead of a tyrosine at position 157, two residues before the second cysteine in strand F. Its ring-NH group forms a hydrogen bond with the main-chain carbonyl of residue 153; this interaction corresponds precisely to the bond formed by the tyrosine-OH group in REI and CD4 domain 1. Another hallmark of immunoglobulin variable domains, the salt bridge between an arginine at the beginning of strand D and an aspartic acid in the EF loop, is partially modified in domain 2. The

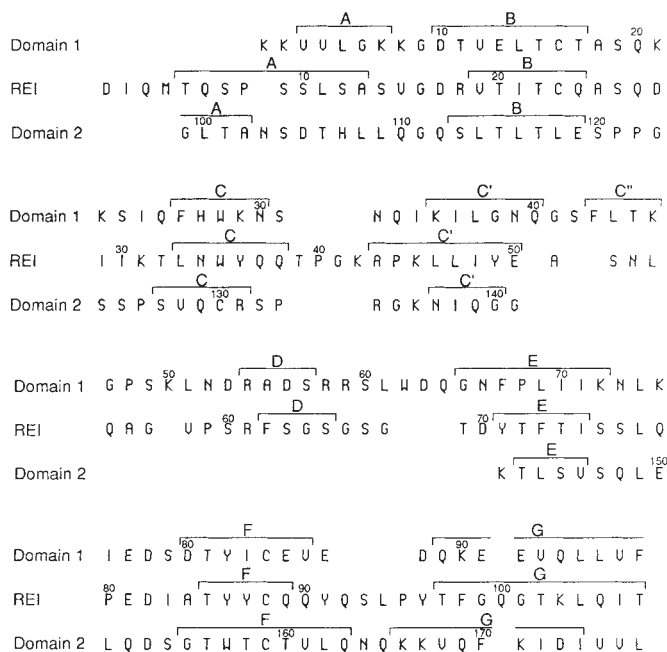


FIG. 4 Sequence alignments implied by the superpositions in Fig. 3. Strands were assigned in CD4 by the method of Kabsch and Sander⁶⁸, using our best-refined coordinate set. Strand assignments in REI are based on the Brookhaven database entry⁵⁵.

TABLE 1 Statistics for data collection, phase determination and refinement

(a) C form							
Crystallization: CD4(1–182), 16–18 mg·ml ^{−1} by hanging-drop method. Well solution: 20% PEG 4K, 0.05 M HEPES (pH 7.85), 0.05 M KCl, 0.1% azide, 20 °C							
Unit cell: Space group C2, <i>a</i> =84.23 Å, <i>b</i> =30.65 Å, <i>c</i> =88.94 Å, β =118.43°							
	Overall	15–6	6–4	Resolution shells (Å)			
				4–3.3	3.3–3	3–2.6	2.6–2.4
Native							
Number of measurements	31,554	2,547	7,485	6,543	4,283	7,423	3,142
Unique reflections	7,829	508	1,249	1,369	1,018	2,180	1,505
Expected unique reflections	8,048	521	1,254	1,372	1,018	2,199	1,684
Completeness (%)	97.7	97.5	99.6	99.8	100	99.6	89.4
Completeness (%) with $I/\sigma > 1$	85.7	97.1	99.2	98.5	95.3	83.9	58.2
R_{merge} (%) [*]	8.6	6.2	8.7	9.0	11.2	15.1	21.1
Mean figure of merit	0.59	0.70	0.68	0.36			
Derivatives							
K_3OsCl_6							
Number of unique reflections	3,046	515	1,240	1,291			
R_{merge} (%) [*]	8.0	7.8	7.5	14.3			
R_c (%) [†]	71	58	83	81			
Phasing power [‡]	1.54	1.19	1.73	1.41			
K_2PtBr_4							
Number of unique reflections	2,913	509	1,202	1,202			
R_{merge} (%) [*]	9.0	8.1	8.7	14.9			
R_c (%) [†]	71	64	69	92			
Phasing power [‡]	1.88	1.48	1.90	1.70			
UO_2SO_4							
Number of unique reflections	1,741	500	1,241				
R_{merge} (%) [*]	7.1	6.1	7.6				
R_c (%) [†]	66	68	62				
Phasing power [‡]	1.10	1.08	1.05				
Refinement:							
R (%) [§] using $F/\sigma > 1$	22.1		16.6	19.9	24.5	28.9	30.5
using $F/\sigma > 3$	21.1		16.6	19.7	23.5	27.1	26.5
(b) A form							
Crystallization: CD4(1–183), 16–18 mg ml ^{−1} by hanging-drop method. Well solution: 20% PEG 4K, 0.05 M HEPES (pH 7.85), 0.1% azide, 4 °C							
Unit cell: Space group C2, <i>a</i> =80.15 Å, <i>b</i> =32.14 Å, <i>c</i> =75.98 Å, β =103.08°							
	Overall	15–6	6–4	4–3.3	3.3–3		
Native							
Number of measurements	7,692	1,729	2,491	2,244	1,228		
Unique reflections	3,400	459	1,039	1,128	774		
Expected unique reflections	3,930	489	1,167	1,290	984		
Completeness (%)	86.5	93.8	89.0	87.4	78.7		
Completeness (%) with $I/\sigma > 1$	72.9	91.6	85.4	73.1	48.5		
R_{merge} (%) [*]	12.5	9.2	9.7	18.3	38.4		
Refinement							
R (%) [§] using $F/\sigma > 1$	19.1		17.1	19.8	23.8		
using $F/\sigma > 3$	18.0		16.7	18.4	21.4		
(c) Refined structure							
R.m.s. deviations of restrained parameters from ideal values							
	Bonds	Angles	Dihedral angles	Chiral and planar restraint	Thermal parameter restraints		
					Bond	Angles	
C form	0.027 Å	4.19°	29.5°	1.97°	0.94 Å ²	1.24 Å ²	
A form	0.025 Å	4.86°	29.3°	1.86°	0.74 Å ²	1.00 Å ²	

Data were collected with CuK α radiation, generated by an Elliot GX-13 rotating anode and collimated by Franks double-mirror optics, using a Siemens-Nicolet area detector controlled by the programs described in ref. 66. Integrated intensities were scaled and merged using the CCP4 program suite, which was also used for many of the subsequent steps in the structure determination. Heavy-atom parameters were refined and MIR phases calculated using the program HEAVY⁶⁷. The three derivatives were prepared in somewhat different ways: (1) K_3OsCl_6 , 1 mM in 25% PEG 4K, 0.05 M HEPES, (pH 7.85) 0.05 M KCl, 3-day soak; (2) K_2PtBr_4 , 1 mM in 25% PEG 4K, 0.05 M HEPES (pH 7.85), with no salt to prevent cell change, 3-day soak and 4–6-hour washing; (3) UO_2SO_4 saturated in 25% PEG 4K, 0.05 M HEPES (pH 7.0), 0.2 M KCl, 2-day soak.

^{*} $R_{\text{merge}} = \sum_h \sum_i |I_{hi} - \bar{I}_h| / \sum_h \bar{I}_h$, where h are unique reflection indices, I_{hi} are intensities of redundant, symmetry-related reflections of index h , and $\bar{I}_h$ is the mean intensity for reflections of index h .

[†] R_c , Cullis R factor for centric reflections.

[‡] Phasing power, mean value of the heavy-atom structure factor amplitudes divided by the residual lack-of-closure error.

[§] $R = \sum |F_o - F_c| / \sum F_o$, where F_o and F_c are observed and calculated structure factor amplitudes, respectively. The refinement does not include reflections at spacings larger than 6 Å.

^{||} For definitions of restraints as applied, see the program manual of XPLOR⁴⁹.

absence of strand D eliminates the usual arginine, but its place seems to be taken by Lys 136 at the CC' corner.

The last few residues of domain 2 appear to be disordered in CD4(1-182). Val 176 is the last residue placed in our most recent model. It lies just at the position where the G strand emerges from the β sheet in the middle of a cluster of hydrophobic residues. The disordered residues 177-182 are likely to form the connection to domain 3. The alignments described above show that they correspond to residues 98-103 (the last of domain 1 and the first five of domain 2). We note that fragments 95-100 and 174-179 are both sequences of hydrophobic residues, suggesting that if the next part of CD4 is indeed immunoglobulin-like in its fold, then the link between domains 2 and 3 may be similar to the one between domains 1 and 2. Many cell-surface proteins are composed of concatenated immunoglobulin-like elements^{25,59}, and the interdomain connection seen in this first view of such a protein is also likely to be found in other structures.

Antibody binding sites

Sites for binding monoclonal antibodies to CD4 have been mapped by several groups by testing their reactivity with CD4 altered at one or more residues^{24,64,65}, by generating 'escape mutants' of CD4²³, or by analysing patterns of cross competition between pairs of antibodies¹⁵. A summary of the results for several monoclonals that map to domain 1 is shown in Fig. 5 by using a diagram of the CD4 structure. Only changes that selectively affect certain monoclonals are shown, as changes that affect a broad spectrum of antibodies probably destabilize the entire domain. As expected, the residues that appear to interact with a given antibody cluster in a single patch on the CD4 surface. Crystallographic analyses of complexes between antibodies and their protein antigens show that the interface is approximately 25-30 Å in diameter⁵⁷. The sites defined by observable changes in binding to mutated CD4 are likely to represent a minimum estimate for the actual antigen-antibody interface, because available variants do not necessarily saturate the surface and because changes in residues near the periphery of the interface may have a less pronounced effect than changes in centrally located side chains. Indeed, most of the patches in Fig. 5 are about 15 Å in diameter. There are clearly two sorts of mutations that affect antibody binding without destroying overall stability: alterations of an exposed residue, defining a contact at the interface, and alterations of a buried residue, probably creating local conformational distortion. Only exposed residues are shown in Fig. 5. An example of a change in a buried residue is A55F, which selectively diminishes binding of OKT4d. The bulky phenylalanine side chain would be expected to distort the C'D corner (residues 47-54), precisely where surface changes affect OKT4d affinity.

The binding sites defined by the summary in Fig. 5 are generally consistent with patterns of cross-blocking between the corresponding antibodies¹⁵. The size of an immunoglobulin antigen-combining region (two domains each the size of domain 1) implies substantial excluded volume, and two antibodies binding anywhere on the same side of domain 1 might be expected to interfere. Thus VIT4 and MT 151, which both 'see' the AGFCC'C' face of domain 1, compete effectively with each other. Neither competes with Leu3a, which seems to bind to the right-hand side of the molecule as depicted in Fig. 5a, nor with OKT4a, which binds to the BED face. T4/18T3A9 competes with both VIT4 and Leu3a, as the mapping of their sites requires. One noninterfering pair consists of Leu3a and OKT4d. The diagram in Fig. 5a suggests that both these antibodies interact with the C'C' edge of domain 1, but that Leu3a approaches from 'above' and OKT4d from 'below'. Monoclonal MT151 selects for escape mutants on both domain 1 (residue 94) and domain 2 (residue 165) (ref. 23). These residues are about 12 Å apart. Mutation of lysine 1, which lies between them, also diminishes MT151 binding⁶⁵.

Class II MHC interaction

Amino-acid substitutions that interfere with functional interaction of CD4 and class II MHC molecules on a target cell occur on both domains 1 and 2 (ref. 38). Mutations in the CC' and C'D corners of domain 1 and along the A strand of domain 2 have the strongest effects (arrows in Fig. 5). Monoclonal OKT4a, which binds to the BED face of domain 1 (Fig. 5), competes with the MHC interaction, as do Leu3a and HIV gp120^{38,39}. Examination of the model shows that there must be a very extended region of lateral contact to account for these interferences. The relative orientation of the two domains produces a noticeable concavity in the overall shape of the two-domain fragment, and this surface might provide a notch for interlocking CD4 with a class II MHC molecule projecting 'downwards' from the opposite cell (in the orientation of Fig. 5). If the shape of the MHC molecule were roughly complementary to this notch, most of the widespread contacts implied by the mutational data could occur. In any case, the distribution of the implied contacts is consistent with the notion that CD4 must project beyond the outer, antigen-presenting domain of the MHC molecule to interact with nonpolymorphic regions near its base. Moreover, the total length of the CD4 molecule is sufficient for CD4 to line up with a T-cell receptor bound to an antigen-presenting class II molecule and to form a specific ternary complex.

Binding of gp120

Mutations that influence gp120 binding cluster near the edge of domain 1 that contains strand C' (refs 23, 24, 34-36, 64, 65). A concordance of such changes is presented in Fig. 6. As in the case of changes that alter antibody binding, these sites include both exposed and buried residues. Mutations that alter affinity of several monoclonal antibodies with nonoverlapping binding sites are not included, because their effect can be interpreted as a general destabilization of domain 1. The mutations in buried

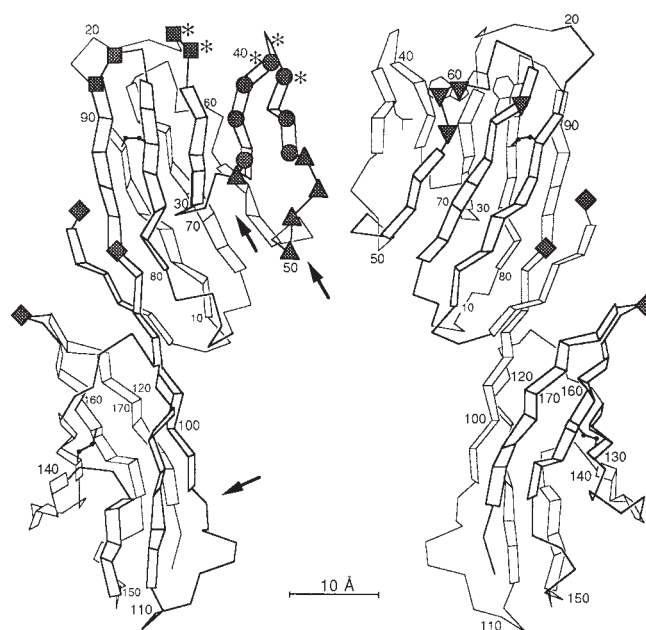


FIG. 5 Monoclonal antibody-binding sites and regions of MHC interaction on domains 1 and 2 of CD4. The directions of view are from the AGFCC'C' face (left) and BED face (right) of domain 1. The left-hand view corresponds roughly to Fig. 2a. The pleated ribbons are β strands; the line segments define connecting loops. Each segment joins successive α -carbon positions. Small numbers label every tenth residue. The symbols represent locations of exposed residue that affect binding of a particular antibody: $\circ$, Leu3a (refs 23, 24, 64, 65); ∇ , OKT4a (refs 23, 24, 64); $\blacktriangle$, OKT4d (refs 23, 24, 64); $\diamond$, VIT4 (refs 23, 24); *, T4/18T3A9 (refs 23, 24); $\blacklozenge$, MT151 (refs 23, 24, 65). Arrows show loops where substitutions affect MHC binding³⁸. In the right-hand view, side chains of Leu 44, Trp 62 and Phe 67 have been included to illustrate the extended hydrophobic core supporting the C'C' loop.

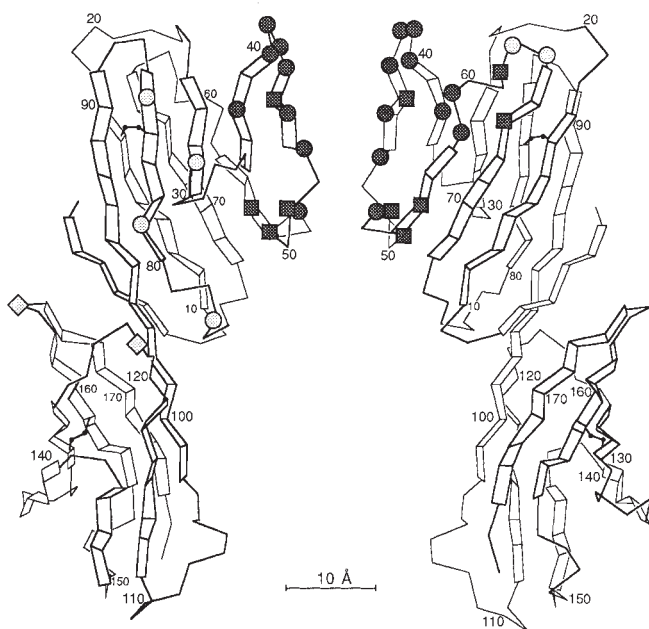


FIG. 6 Residues that affect gp120 interaction with CD4. The views are described in the caption to Fig. 5. The symbols represent classes of residues as follows: ●, exposed (refs 23, 24, 36, 64, 65); ■, buried (refs 24, 34, 36, 64); ⊗, some exposed residues that are displaced from the C'C' ridge (ref. 65); ⊕, positions in domain 2 reported to have some effect on gp120 interaction (refs 34, 35).

residues shown in Fig. 6 can therefore be presumed to influence binding by generating local, rather than global, perturbations in the folded structure. Many of the exposed residues that can be inferred to contact gp120 lie between positions 38 and 52. Their side chains decorate the C' edge, forming a ridge about

25 Å long (see Fig. 2b). They include an exposed phenylalanine at position 43 in the C'C' turn. Several mutations from the adjacent β -sheet surfaces also affect the gp120 interaction. Several of them involve side chains, such as Arg 59, that project toward the ridge, but results of a recent survey by alanine-scanning mutagenesis⁶⁵ implicate some residues displaced farther from the C' edge (see Fig. 6). Because changes in the A and B strands at the opposite edge of the domain and in the BC loop at the 'top' do not affect gp120 binding, we propose that the viral glycoprotein approaches domain 1 from the C' side and that it may fit over the two faces to some extent. It appears not to contact the surface of the CC' corner (residues 30–35), which projects strongly from the AGFCC'C' face. The C' ridge is supported at either end by hydrophobic pockets, in which mutations of buried residues reduce gp120 affinity. Examples of such mutations are W62C and F67L, which would perturb packing of side chains in the extended hydrophobic core beneath the C'C' turn, and S49Y and A55F, which would perturb the C'D corner.

What can we infer about the complementary site on gp120? Definition of the contact surface on CD4 by mutagenesis is clearly still incomplete, but the summary in Fig. 6 shows that it is likely to require a region of interaction on gp120 at least 25 Å long and 12 Å wide. The data suggest that the gp120 site is a groove rather than a flat surface, but specifically designed mutations are needed to confirm this and to estimate a depth. Because gp120 is a large protein, there may be additional contacts. These might be less critical for binding, and they might therefore not be detected by competition experiments with truncated forms of CD4. This possibility could explain reports of effects from changes near residues 122 and 164 in domain 2^{34,35} and from a substitution at position 77 (ref. 65). If these observations do reflect local interactions, some part of gp120 could be imagined to extend across the middle of CD4 as seen in Fig. 6. The C' ridge seems nonetheless to be the most important surface from the perspective of inhibitor design. □

Received 16 October; accepted 23 October 1990.

- Reinherz, E. L., Kung, P. C., Goldstein, G. & Schlossman, S. F. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4061–4065 (1979).
- Reinherz, E. L., Kung, P. C., Goldstein, G. & Schlossman, S. F. *J. Immun.* **234**, 2894–2896 (1979).
- Meuer, C., Schlossman, S. F. & Reinherz, E. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4395–4399 (1982).
- Biddison, W., Rao, P., Thallie, M., Goldstein, G. & Shaw, S. *J. exp. Med.* **156**, 1065–1076 (1982).
- Krensky, A. M., Reiss, C. S., Mier, J. W., Strominger, J. L. & Burakoff, S. J. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2365–2369 (1982).
- Gay, D. *et al. Nature* **328**, 626–629 (1987).
- Engleman, E. C., Benike, C. J., Grumet, C. & Evans, R. L. *J. Immun.* **127**, 2124–2129 (1981).
- Doyle, C. & Strominger, J. L. *Nature* **330**, 256–259 (1987).
- Kruisbeek, A. M. *et al. J. exp. Med.* **161**, 1029–1047 (1985).
- Eichmann, K., Jonsson, J. I., Falk, I. & Emmrich, F. *Eur. J. Immun.* **17**, 643–650 (1987).
- Rudd, C. E., Trevillyan, J. M., Dasgupta, J. D., Wong, L. L. & Schlossman, S. F. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5190–5194 (1988).
- Veillette, A., Bookman, M. A., Horak, E. M. & Bolen, J. B. *Cell* **56**, 301–308 (1988).
- Klatzmann, D. *et al. Science* **225**, 59–63 (1984).
- Dalgleish, A. *et al. Nature* **312**, 763–766 (1984).
- Sattentau, Q., Dalgleish, A., Weiss, R. & Beverley, P. C. L. *Science* **234**, 1120–1123 (1986).
- McDougal, J. S. *et al. J. Immun.* **137**, 2937–2944 (1986).
- McDougal, J. S. *et al. Science* **231**, 382–385 (1986).
- Maddon, P. *et al. Cell* **47**, 333–348 (1986).
- Sodroski, J., Goh, W. C., Rosen, C., Campbell, K. & Haseltine, W. A. *Nature* **322**, 470–474 (1986).
- Lifson, J. *et al. Nature* **323**, 725–728 (1986).
- Maddon, P. *et al. Cell* **42**, 93–104 (1985).
- Clark, S. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 1649–1653 (1987).
- Peterson, A. & Seed, B. *Cell* **54**, 65–72 (1988).
- Sattentau, Q. *et al. J. exp. Med.* **170**, 1319–1334 (1989).
- Williams, A. F., Davis, S. J., He, Q. & Barclay, A. N. *Cold Spring Harb. Symp. quant. Biol.* **54**, 637–647 (1989).
- Sayre, P. & Reinherz, E. L. *Eur. J. Immun.* **15**, 291–295 (1985).
- Smith, D. H. *et al. Science* **238**, 1704–1707 (1987).
- Fisher, R. A. *et al. Nature* **331**, 76–78 (1988).
- Hussey, R. E. *et al. Nature* **331**, 78–81 (1988).
- Deen, K. C. *et al. Nature* **331**, 82–84 (1988).
- Trautwein, A., Luke, W. & Karjalainen, K. *Nature* **331**, 84–86 (1988).
- Berger, E. A., Fuerst, T. R. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2357–2361 (1988).
- Richardson, N. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 6102–6106 (1988).
- Clayton, L. K. *et al. Nature* **335**, 363–366 (1988).
- Mizukami, T., Fuerst, T. R., Berger, E. A. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **85**, 9273–9277 (1988).
- Arthos, J. *et al. Cell* **57**, 469–481 (1989).
- Landau, N. R., Warton, M. & Littman, D. R. *Nature* **334**, 159–162 (1988).

- Clayton, L. K., Sieh, M., Pious, D. A. & Reinherz, E. L. *Nature* **339**, 548–551 (1989).
- Lamarre, D. *et al. Science* **245**, 743–746 (1989).
- Davis, S. J. *et al. J. molec. Biol.* **213**, 7–10 (1990).
- Kwong, P. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 6423–6427 (1990).
- Harrison, S. C. *et al. manuscript in preparation*.
- Ryu, S.-E. *et al. Nature* **348**, 419–426 (1990).
- Garlick, R. L. *et al. AIDS Res. hum. Retrovirus* **6**, 465–479 (1990).
- Wang, B. C. *Meth. Enzym.* **115**, 90–111 (1985).
- Bricogne, G. *Acta crystallogr. A* **32**, 832–847 (1976).
- Jones, T. A. *J. Appl. crystallogr.* **11**, 268–272 (1978).
- Brünger, A., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
- Brünger, A. T. *XPLOR Manual, Version 2.1*. (Yale University, New Haven, Connecticut, USA).
- Stuart, D. & Artymuk, P. *Acta crystallogr. A* **40**, 713–716 (1985).
- Crowther, R. A. in *The Molecular Replacement Method* 173–183 (ed. Rossmann, M. G.) (Gordon and Breach, New York, 1972).
- Huber, R. in *Molecular Replacement. Proceedings of the Daresbury Study Weekend* 58–61 (SERC, Daresbury, UK, 1985).
- Sussman, J. L., Holbrook, S. R., Church, G. M. & King, S. H. *Acta crystallogr.* **33**, 800–804 (1977).
- Bjorkman, P. J. *et al. Nature* **329**, 506–511 (1987).
- Epp, O. *et al. Biochemistry* **14**, 4943–4952 (1975).
- Lesk, A. M. & Chothia, C. *J. molec. Biol.* **160**, 325–342 (1982).
- Colman, P. M. *Adv. Immun.* **43**, 99–132 (1988).
- Kabat, E. *et al. Sequences of Proteins of Immunological Interest* (US Dept. of Health & Human Services, Bethesda, Maryland, 1987).
- Hunkapiller, T. & Hood, L. *Adv. Immun.* **44**, 1–63 (1989).
- Tramontano, A., Chothia, C. & Lesk, A. M. *J. molec. Biol.* **215**, 175–182 (1990).
- Richardson, J. & Richardson, D. C. in *Prediction of Protein Structure and the Principles of Protein Conformation* (ed. Fasman, G.) (Plenum, New York, 1989).
- Kirsbaum, L. *et al. J. Immun.* **142**, 3931–3936 (1989).
- Holmgren, A. & Brändén, C. I. *Nature* **342**, 248–251 (1989).
- Brodsky, M. H., Warton, M., Myers, R. M. & Littman, D. R. *J. Immun.* **144**, 3078–3086 (1990).
- Ashkenazi, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 7150–7154 (1990).
- Blum, M., Metcalf, P., Harrison, S. C. & Wiley, D. C. *J. appl. Crystallogr.* **20**, 235–242 (1987).
- Terwilliger, T. C. & Eisenberg, D. *Acta crystallogr. A* **39**, 813–817 (1983).
- Kabsch, W. & Sander, C. *Biopolymers* **22**, 2577–2637 (1983).

ACKNOWLEDGEMENTS. We thank D. C. Wiley and many members of our laboratories for discussion and advice; N. B. Jansen, H. F. Meyer, D. P. Palermo, D. P. Wells and others at the Upjohn Co. for preparation of CD4 fragments; W. Weiranga, L. Post and H. Einspahr, also of Upjohn, for support and encouragement; W. Kabsch and K. C. Holmes, for advice on phase combination; W. Hendrickson and colleagues for exchanging manuscripts before submission; and R. Hellmiss for help with illustrations. The work was supported by the NIH (grants to S.C.H., E.L.R. and D. C. Wiley). Atomic coordinates have been deposited in the Brookhaven database, accession number 2CD4.

Crystal structure of an HIV-binding recombinant fragment of human CD4

Seong-Eon Ryu*, Peter D. Kwong*, Alemseged Truneh†, Terence G. Porter†, James Arthos†, Martin Rosenberg†, Xiaoping Dai‡, Nguyen-huu Xuong‡, Richard Axel*§, Raymond W. Sweet† & Wayne A. Hendrickson*§

* Department of Biochemistry and Molecular Biophysics and § Howard Hughes Medical Institute, Columbia University, New York, New York 10032, USA

† SmithKline Beecham Pharmaceuticals, King of Prussia, Pennsylvania 19406, USA

‡ Departments of Physics, Chemistry and Biology, University of California at San Diego, La Jolla, California 90293, USA

CD4 glycoprotein on the surface of T cells helps in the immune response and is the receptor for HIV infection. The structure of a soluble fragment of CD4 determined at 2.3 Å resolution reveals that the molecule has two intimately associated immunoglobulin-like domains. Residues implicated in HIV recognition by analysis of mutants and antibody binding are salient features in domain D1. Domain D2 is distinguished by a variation on the β -strand topologies of antibody domains and by an intra-sheet disulphide bridge.

CD4, a cell-surface glycoprotein found primarily on T lymphocytes, is required to shape the T-cell repertoire during thymic development and to permit appropriate activation of mature T cells¹. T cells that recognize antigens associated with class II major histocompatibility complex (MHC) molecules, mainly T helper cells, express CD4. Evidence is accumulating that CD4 and the T-cell receptor coordinately engage class II molecules on antigen-presenting cells to mediate an efficient cellular immune response, and that engaged CD4 may transmit a signal to an associated cytoplasmic tyrosine kinase, p56^{lck}.

CD4 belongs to the immunoglobulin superfamily of molecules which generally serve in recognition processes^{2,3}. The sequence of CD4^{4,5} indicates that it consists of a large (~370 residues) extracellular segment composed of four tandem immunoglobulin-like domains, a single transmembrane span, and a short (38 residues) C-terminal cytoplasmic tail. The first domain (D1) shares several features with immunoglobulin variable domains, but the sequence similarities between immunoglobulins and the other extracellular domains (D2, D3 and D4) are more remote.

In humans, CD4 can be subverted from its normal immunosupportive role to become the receptor for infection by the human immunodeficiency virus (HIV)^{1,6,7}. Recombinant soluble CD4 proteins bind to the HIV envelope glycoprotein gp120, and can thus inhibit viral infection and virus-mediated cell fusion *in vitro* (refs 8, 9 and references therein). Domain D1 suffices for high-affinity binding to gp120 (ref. 8), and the analysis of substitution mutants further limits the sites of interaction to discrete regions in the domain^{8,10-18}.

Crystals of whole soluble CD4 (sCD4) molecules have been grown^{19,20} but there has been limited success in achieving adequate diffraction order. The high solvent content and weak diffraction of several characterized polymorphs of human sCD4 are compatible with an extended, flexible molecule¹⁹. From the pattern of proteolytic cleavages that generate stable fragments (refs 21-23 and unpublished results), the main flexibility seems to be at the D2 to D3 junction. We have now crystallized a truncated derivative of CD4 that diffracts well, and here we report its atomic structure. This recombinant fragment⁸ as

secreted from Chinese hamster ovary (CHO) cells consists of residues 1-183 of human CD4 plus two missense residues, Asp-Thr; and it is unglycosylated. This molecule, which we refer to as D1D2, is as active as sCD4 in binding to gp120 (dissociation constant $K_d \approx 3$ nM) and retains all antibody epitopes mapped to these domains of CD4 (ref. 8 and unpublished results). Others have crystallized similar fragments from the N-terminal half of sCD4^{24,25} and the structure of one is reported in the accompanying paper²⁵.

Here we describe the D1D2 structure in comparison with that of immunoglobulin domains, provide a geometrical definition for HIV recognition sites, and discuss implications of the structure for normal CD4 function and evolution of the immunoglobulin family. We find that the domains of CD4 are indeed immunoglobulin-like, although there are significant differences from the antibody analogues. The primary sites for HIV interaction are on loops that protrude from the variable-like D1 domain in analogy with immunoglobulin complementarity-determining regions (CDRs). The D2 domain, which is intimately associated with D1, resembles constant domains but it is distinguished by a strand topology that is variable-like.

Structure determination

D1D2 protein secreted from CHO cells was purified to homogeneity by following procedures similar to those used for sCD4 (ref. 26). Crystals of this protein grown from polyethylene glycol (PEG) and stabilized at pH 8.2 (Table 1) belong to space group C2 and have unit cell dimensions of $a = 83.71$ Å, $b = 30.07$ Å, $c = 87.54$ Å, $\beta = 117.3^\circ$. They contain one D1D2 molecule per asymmetric unit and 50% solvent. In searching for derivatives, we soaked crystals into stabilization medium doped with various heavy-atom compounds.

Diffraction data for the structure determination (Table 1) were measured at an area detector facility²⁷ where we collected CuK α data for the native protein and for candidate derivatives. Observable intensities could be recorded to 1.9 Å spacings from fresh native crystals, and even more strongly from the K₂Pt(NO₂)₄ derivative. We could also measure multiwavelength anomalous diffraction (MAD) data from this platinum derivative using characteristic gold emission lines which bracket the PtL_{III} absorption edge²⁸. Phases were evaluated by both multiple isomorphous replacement (MIR) and MAD methods (Table 1). Electron density maps from MIR phasing (including anomalous data) at 2.7 Å resolution and from MAD phasing at 3.5 Å resolution both showed similar features. We then probabilistically combined MIR and MAD phase information to produce a combined map at 2.7 Å resolution (Fig. 1a) which showed distinct solvent channels and an apparent two-domain structure. A complete chain-tracing consistent with the amino-acid sequence for the first domain was possible, but the second domain remained difficult to interpret. By using a hand-drawn molecular envelope, we then performed solvent flattening and density truncation to improve the map further (Fig. 1b). This enabled us to trace the chain through the second domain. The C-terminal 12 residues (174-185) could not be seen.

A complete model was built into the solvent-flattened map as displayed by FRODO²⁹ program on a graphic workstation. Fragments from well-refined structures were fitted onto Ca guide points measured from a minimap and used as the starting point for building. Refinement made use of PROLSQ³⁰ and XPLOR³¹ programs and included several manual rebuildings. Resolution of the analysis was gradually extended to 2.3 Å. After analysis of model 5 (Table 1) at $R = 0.208$, we discovered on comparing manuscripts that this model and one developed

by Wang *et al.*²⁵ differed in the alignments of sequence in two strands (E of D1 and B of D2) and in conformations at residues 1, 103–108, 134–139 and 151–154. We then tested a model with residues 66–73 and 112–120 positioned in the former places of 64–71 and 114–122, respectively, and with the 134–139 region also rebuilt. Further refinement produced model 6 (Table 1) with an R value of 0.196 and stereochemical ideality typified by an r.m.s. deviation of 0.018 Å from ideal bond lengths. Study of this still incompletely refined model confirms the revised

TABLE 1 Statistics from the crystallographic analysis

(a) Diffraction data									
Derivative	$d_{\min}$ (Å)	X-ray line /wavelength (Å)		No. of measurements	No. of reflections		Data coverage (%)		R_{merge}
Native	2.3	CuK α /1.542		24,407	7,882		92		0.061
K ₂ Pt(NO ₂) ₄	2.5	CuK α /1.542		40,130	6,757 (12,667)		99 (93)		0.057 (0.053)
	2.7	AuL β_1 /1.084		33,914	5,314 (9,978)		97 (97)		0.069 (0.062)
	2.5	AuL β_2 /1.068		43,005	6,639 (12,505)		97 (92)		0.072 (0.065)
	2.7	AuL α_1 /1.276		30,724	5,432 (10,210)		99 (99)		0.062 (0.054)
K ₂ PtCl ₆	2.8	CuK α /1.542		22,033	4,786 (8,751)		96 (88)		0.094 (0.086)
(b) MIR analysis									
$d_{\min}$ (Å)	11.1	7.7	5.9	4.8	4.0	3.4	3.0	2.7	Total
K ₂ Pt(NO ₂) ₄									
r.m.s. $f_{\text{H}}/E_{\text{iso}}$	2.18	1.69	2.01	1.48	1.04	1.12	1.24	1.29	1.32
r.m.s. $\Delta F_{\text{calc}}/E_{\text{ano}}$	0.85	0.88	1.08	1.08	0.79	0.74	0.49	0.37	0.63
K ₂ PtCl ₆									
r.m.s. $f_{\text{H}}/E_{\text{iso}}$	1.54	1.72	1.92	1.52	1.13	1.19	1.23	—	1.35
r.m.s. $\Delta F_{\text{calc}}/E_{\text{ano}}$	0.45	0.54	0.58	0.39	0.24	0.21	0.14	—	0.23
$\overline{m}$	0.81	0.71	0.75	0.71	0.58	0.57	0.51	0.37	0.53
(c) MAD analysis									
Observed ratio					(d) Solvent flattening				
λ (Å)					Shell/cycle		R	$\Delta\phi$	
	1.542	1.084	1.068	1.276	1/1		0.31	28.2	
					1/5		0.24	35.5	
1.542	0.043	0.066	0.057	0.069	3/5		0.23	37.0	
1.084		0.036	0.029	0.030	7/5		0.22	43.8	
1.068			0.038	0.036	13/10		0.21	63.4	
1.276				0.037					
(e) Refinement									
Step	Program	Data range (Å)		B -value refinement	Rebuilding	H ₂ O	R	Δd (Å)	
0	PROLSQ	10.0–2.7		No	No	0	0.447	0.044	
1	PROLSQ	5.0–2.7		No	No	0	0.271	0.025	
2	PROLSQ	10.0–2.3		Yes	Yes	0	0.273	0.024	
3	XPLOR	10.0–2.3		No	Yes	0	0.268	0.024	
4	PROLSQ	10.0–2.3		Yes	Yes	0	0.250	0.023	
5	PROLSQ	10.0–2.3		Yes	Yes	106	0.208	0.020	
6	XPLOR/PROLSQ	10.0–2.3		Yes	Yes	72	0.196	0.018	

Crystallization. D1D2 was purified in sequential steps of ion-exchange chromatography²⁶ involving S-Sepharose capture at pH 5.2, Q-Sepharose passage at pH 9.0 and a final elution from S-Sepharose at pH 6.0 in 0.6 M NaCl. Crystallization was carried out at 20 °C by equilibration against reservoirs of 20–25% PEG 3350 starting from hanging droplets composed of protein at ~20 mg ml⁻¹ in PBS at pH 7.0 and equal volumes from the reservoir. Collected crystals were transferred to a stabilization medium (30% PEG 3350, 80 mM Tris buffer, pH 8.2). Typical Crystals have dimensions of 0.5 × 0.3 × 0.2 mm. **Diffraction data** for MIR phasing and for refinement were measured with CuK α X rays generated by a Rigaku RU200 rotating anode and detected on two multiwire chambers on the Mark III system at UCSD. All data were collected at or near 20 °C. The inverse beam method was used to collect Bijvoet mates; that is, settings were changed from (ω, χ, ϕ) to $(\omega, -\chi, \phi + \pi)$ after rotations at each crystal orientation. Data for MAD phasing were collected on the Mark II area detector system. A graphite monochromator was used to select specific characteristic lines produced from a gold-plated anode installed on an Elliot GX-6 generator. The monochromator was tuned to 50% intensity at one side or the other of the unresolved L β_1 /L β_2 doublet for selection of a primary monochromatic component. The UCSD area detector processing package was used for the reduction of all data. Values given inside of parentheses for numbers of reflections and R_{merge} are without merging of Friedel mates ($R_{\text{merge}} = \sum |I_{\text{obs}} - I_{\text{avg}}| / \sum I_{\text{avg}}$). **MIR analysis.** Heavy-atom positions were determined from Patterson syntheses. The refinement of heavy-atom parameters and the phase calculations were performed with a version of program PHARE modified by Z. Otwinowski (personal communication) for an improved error treatment. Relative occupancies for the K₂Pt(NO₂)₄ and K₂PtCl₆ derivatives were 1.00, 0.94) and (0.53, 0.51, 0.25, 0.22, 0.19), respectively. The first two heavy-atom sites of the K₂PtCl₆ derivative are in common with the sites of the K₂Pt(NO₂)₄ derivative. Parameters cited in the table are: f_H , calculated heavy-atom structure-factor contributions; ΔF_{calc} , calculated Bijvoet difference; E_{iso} , r.m.s. isomorphous closure error; E_{ano} , r.m.s. anomalous closure error; and $\bar{m}$, mean figure-of-merit. **MAD analysis.** The MADSYS system of programs⁵⁴ was used to produce initial MAD phases. Phase combination between MIR and MAD was performed by the combination of ABCD coefficients⁵⁵. But, in this case it was necessary to modify the original MAD refinement formulation⁵⁴ so as to produce phases for the native non-anomalous scattering component⁵⁶ of the platiny CD4 complex. The previously used method for extracting ABCD coefficients⁵⁵ proved to be ineffective in this case; instead, these phasing coefficients were extracted directly from MAD phases and figures-of-merit⁵⁷. MIR and MAD phases were combined at 20.0–3.5 Å; MIR phases were used alone for the 3.5–2.7 Å data. Observed anomalous diffraction difference ratios are given for the data between 20.0–3.5 Å. Bijvoet (diagonal elements of matrix) and dispersive (off-diagonal elements) difference ratios represent r.m.s. ($\Delta F_{\text{H}}/r.m.s. (F)$) and r.m.s. ($\Delta F_{\text{A}}/r.m.s. (F)$), respectively. **Solvent flattening** was performed following procedures used in the structure analysis of myohemerythrin⁵⁸. After interpretation of the first domain, a molecular envelope was delineated which included 58% of the asymmetric unit of the crystal (as compared with 50% for the true solvent content) to ensure that all protein density was included. After subtraction of the mean solvent density and truncation of the lowest density values within the molecular boundary, the contents of the modified density function were Fourier-inverted. Phasing coefficients from the inversion were reduced by 0.5 before combination with ABCD values from MIR. A truncation level to eliminate the lowest 20% of molecular points (as compared with 0% and 10%) was chosen as giving the best definition for known features in domain D1. The phase refinement was carried out starting from the 20.0–4.5 Å shell, and including higher angle reflections successively until finally reaching 2.7 Å spacings. Five cycles of solvent flattening were carried out for each shell between shell 1 and shell 12, and 10 cycles were run for shell 13. R , Crystallographic R value between observed structure factor magnitudes and calculated values from the solvent-flattened map. $\Delta\phi$, Phase differences between phases from MIR plus MAD and phases from the solvent flattened map. **Refinement.** Selected points along the course of refinement are listed. For all steps, the data having $F_{\text{obs}} > 4\sigma(F_{\text{obs}})$ were included. Step 0 refers to the unrefined starting model after adjustment of the scale factor only. From step 2 onward, the data between 10.0 and 5.0 Å were included which helped to achieve better connectivities in $2F_{\text{obs}} - F_{\text{calc}}$ maps. The final model at 2.3 Å resolution is based on the 5,689 reflections (64%) which met the 4 σ criterion. This partially refined model includes the 1,420 atoms from residues 1–173 plus the 72 water sites. Thermal parameters were restrained as typified by an r.m.s. discrepancy of 1.17 Å in B values between bonded main-chain atoms. $R = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$; Δd , r.m.s. bond deviation from ideality (Å).

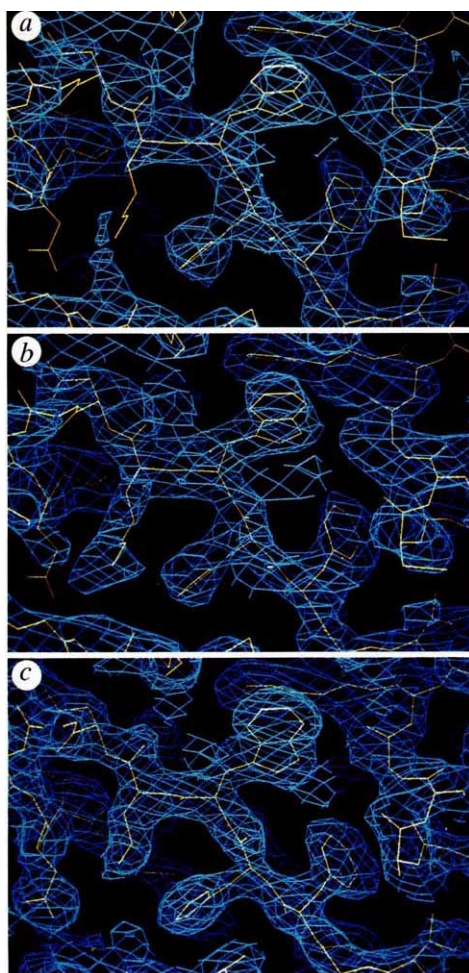


FIG. 1 Electron-density distributions used in the structural determination. The portion displayed in each panel includes the segment Phe 26-His 27-Trp 28-Lys 29 with the partially refined model 5 (yellow) superimposed on the density (blue). *a*, Experimental map at 2.7 Å resolution based on combined MIR plus MAD phase information; $\overline{m}=0.58$. *b*, Experimental map at 2.7 Å resolution after phase refinement by solvent flattening and density truncation; $\overline{m}=0.86$. *c*, Refined $2|F_{\text{obs}}|-|F_{\text{calc}}|$ map with model phases after refinement at 2.3 Å resolution; $R=0.208$.

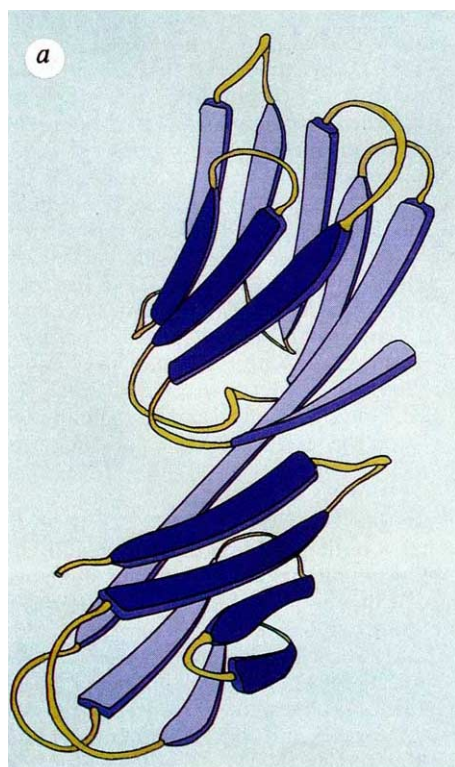
alignment in D1 and supports the changes made in D2, although several loops in D2 remain ill-defined. All conclusions drawn from model 5 remain unchanged, but the precise numerical quantities and Figs 2*b*, 4 and 5*a* were redone from model 6.

Overall structure

The D1D2 structure is of the all- β type. It consists of two intimately associated domains, each of which is a β sandwich folded with immunoglobulin-strand topology. The polypeptide folding is shown in Fig. 2. The first domain (D1) comprises residues 1–98 disposed in nine β strands, and the second domain (D2) contains residues 99–173 and has seven β strands. These domain boundaries are in striking correspondence with intron boundaries immediately after residues 100 and 177 in the gene structure⁵. The suggestion that J-like regions follow these introns is not borne out by the structure.

The last strand of D1 continues straight into the first strand of D2, running for a length of 49 Å from residues 88–103. This tandem association of domains leads to a rod-shaped molecule of dimensions roughly $25 \times 25 \times 60$ Å. By comparing the solvent-accessible surface area of D1D2 with that of the separated

FIG. 2 Backbone structure of the D1D2 fragment of CD4. *a*, Schematic diagram (copyright Yarmolinski and Hendrickson, 1990). *b*, Stereodigram of the α -carbon backbone. Positions of residue numbers divisible by ten are indicated by small spheres, sulphur atoms in disulphide bridges are indicated by larger spheres. The point of view is down the Z' axis after rotations about $X(100^\circ)$, $Z'(-30^\circ)$, $Y'(-90^\circ)$ and $X''(10^\circ)$ starting with the frame having X, Y, Z along *a, b* and *c**.



domains, we found that 310 Å² from each domain is buried in the interface. This compares with 110 Å² for the VH-CH interface of Fab New³². Interdomain contacts are largely hydrophobic. Residues involved in interdomain contacts include Val 3, Leu 5, Ile 76, Leu 96, Val 97 and Phe 98 from D1, and Glu 119, Pro 121, Gln 163, Val 168 and Phe 170 from D2. The surface of D1D2 has many positively charged amino acids (the calculated pI of residues 1–183 is 10.0), and the electrostatic potential surface³³ computed from the model is mostly positive at neutral pH. There are, however, two prominent patches of negative potential: one is associated with Glu 85, Glu 87 and Asp 88 and the other is from Asp 105, Asp 153 and Asp 173.

TABLE 2 Residues implicated in HIV recognition by point mutation in D1 of CD4*

Amino-acid residue	Fractional solvent accessibility†	Effect on gp120 binding‡§		Effect on syncytium formation‡	Ref.
K29	0.30	A	■□□		16
G38	(0.31)	E	□□□	E ■□□	10, 15
Q40¶	0.85	P	□□□	H □□□	10, 12
		H	□□□	H □□□	8, 15
		A	■□□		16
G41	(0.93)	S	■□□		15
S42	1.00	Y	■□□	A □□□	15, 16
F43	0.90	V	■□□	L ■□□	8, 15
		L	□□□	L □□□	10
L44	0.38	S	■□□	A ■□□	15, 16
T45	0.69	P	■□□	G ■□□	10, 15
		I	□□□	A □□□	8, 16
G47	(0.43)	E	□□□	S ■□□	10, 15
		R	■□□	R ■□□	10
S49	0.58	T	□□□	Y ■□□	15
		A	■□□		16
N52	0.71	K	□□□	A ■□□	15, 16
A55	0.00	F	■□□		8
R58	0.73	G	■□□	A □□□	15, 16
R59	0.82	G	■□□	A ■□□	15, 16
Q64	0.84	E	□□□	A ■□□	15, 16
		K	□□□		12
F67	0.00	L	■□□		15
E77	0.97	A	■□□		16
T81	0.55	A	■□□		16
E85	0.59	A	■□□		16
E87	0.53	K	□□□	K ■□□□	10, 17
		A	□□□	G ■□□□	16, 17
D88	0.71	A	□□□	N ■□□□	16, 17
Q89	0.89	L	□□□	L ■□□□	10, 17
		A	□□□	K ■□□□	16, 17

* Within the 98 amino-acid residues of the D1 domain, there is no mutational information on residues 3–6, 12, 14, 16, 18, 26, 69–71, 76, 83, 84 and 97. However, the binding of gp120 to Rhesus CD4¹⁰ and to the human-rat chimeric protein eliminates 12, 14, 18 and 76 from this list. Mutations which caused global alteration of structure, as evidenced by reduced binding of several anti-CD4 monoclonal antibodies, were not considered^{15,16}. Such mutations were the only data for 13 amino acids and thus led to exclusion of residues 7, 13, 28, 36, 37, 54, 57, 62, 78, 79, 82, 93 and 95. Thus, in total, 25 amino acids are excluded from consideration.

† Fractional solvent accessibility was calculated as the ratio of solvent-accessible surface area for atoms of an amino-acid residue X in the protein to that area obtained after reducing the structure to a Gly-X-Gly tripeptide⁵⁹. Values cited are for side-chain atoms except in the case of glycine residues where main-chain values are given.

‡ Mutants are identified by single-letter code. Mutational effects are symbolized by blackening in steps proportional to the quantitative sensitivity of the assays as reported. Only single amino-acid substitutions are included. The results from double substitutions largely confirm these data. Certain double mutations are of particular note: A, G41S/F43C severely disrupts binding¹⁵, emphasizing the significance of F43; B, L51M/A55S does not affect binding¹⁵, indicating that the severe effect by the A55F substitution is conformational in nature; C, E87G/K90G¹⁵ and D88N/Q89R⁸ do not affect binding, further indicating that this region is not involved in initial binding to gp120.

§ Relative effects of the indicated mutations on virus¹⁰ or gp120¹⁵ binding to cell surface CD4 or gp120 binding to soluble CD4 proteins^{8,12,16}. The results shown are an interpretation of the primary data and comparisons among the results from the different laboratories are approximate.

|| Relative ability of cells bearing CD4 mutant proteins to form syncytium with cells expressing viral envelope proteins^{10,17} or of soluble CD4 proteins to inhibit syncytium⁸.

¶ The Q40A mutation increased binding affinity to gp120 twofold.

The D1D2 crystal lattice contains diad axes that could accommodate a dimeric molecule. However, contacts across screw axes dominate the lattice interactions and the only diad-mediated interface (between D1 domains) seems unlikely to persist in solution. The lattice interactions of D1 are more extensive than those of D2 which is reflected in molecular mobilities of the domains: the average *B* value is 19.9 Å² for D1 compared with 37.8 Å² for D2. This accounts for our greater difficulty in tracing the D2 chain.

CD4, like many members of the immunoglobulin superfamily, is a single-chain cell-surface receptor composed of tandemly repeated domains. Some of these domains have been assigned as V-like and others as C-like or in the 'C2 set'³⁴. The D1D2 domain of CD4 contains both types and, apart from the similarity of CD5 with the PapD bacterial chaperone protein³⁵, this is the first three-dimensional structure for such members of the family.

Variable-like domain D1

As anticipated by sequence comparisons, D1 is similar to the variable (V) domains of antibodies. A schematic drawing of the folding pattern with standard immunoglobulin designations for the secondary structural elements is shown in Fig. 3. Strands B, D and E make up one β sheet, and strands A, C, C', C'', F and G are in the other sheet of the β sandwich. In immunoglobulin V domains, the first half of strand A is hydrogen-bonded to strand B of the outer sheet but the second half switches to the inner sheet. Strand A of D1 is foreshortened and occupies only the 'inner-sheet' position, hydrogen-bonded in parallel with strand G. D1 preserves the inter-sheet disulphide bridge and several other elements of the hydrophobic core characteristic of the immunoglobulin framework. Starting from these alignments we have superimposed D1 onto representative immunoglobulin V domains from Bence-Jones protein Rei (VLκ)³⁶, and Fab New (VLA and VH)³². The structurally aligned sequences are shown in Fig. 4. D1 superimposes remarkably well onto the β-sheet framework of all V domains, with a best match to Rei bringing 72 Cα atoms to an r.m.s. discrepancy of 1.22 Å from exact superposition.

In striking contrast with the conserved core of β strands, several loops between strands in CD4 are quite different from those in immunoglobulin V domains. In particular, loop CC' is shortened by four residues and loop FG (immunoglobulin CDR3) is shortened by four to six residues from canonical immunoglobulin lengths. These loops mediate VH-VL dimerization in immunoglobulins, an interaction that does not occur with CD4. One addition, important for gp120 binding, is the lengthening of loop C'C'' (immunoglobulin CDR2) by three residues in CD4 compared with κ light chains such as Rei. In this regard, CD4 more closely resembles VH domains even though overall it superposes best with Rei. This CDR2-like loop interacts with Trp 62, not found in antibodies, and juts out at the tip. Few of the loop changes in D1 were detected in previous alignments with Rei, and consequently structural predictions based on Rei³⁷ have failed to capture the essence of CD4. The salience of the CDR2 analogue of CD4 and the diminutive nature of the CC' loop and the CDR3 analogue are evident in Figs 2 and 3.

Distinctive domain D2

Domain D2 of CD4 has a tertiary folding that can readily be recognized as resembling immunoglobulin constant domains. Again following standard immunoglobulin nomenclature (Fig. 3), one β sheet of the sandwich-like structure contains strands A, B and E, and the other sheet has strands C, C', F and G. Despite the general similarity of D2 to constant domains, details of the folding are considerably divergent. First, the size of D2 is small (75 residues) compared with that of immunoglobulin constant domains (~100 residues). This manifests itself in shorter strand lengths. A second variation is that strand C',

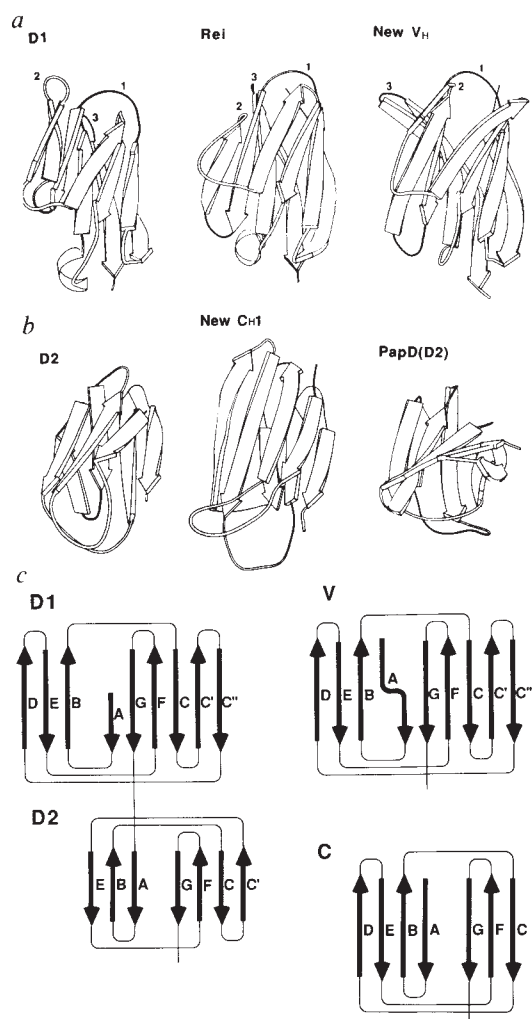


FIG. 3 Schematic comparisons between domains of CD4 and other immunoglobulin-folded domains. *a*, Ribbon diagrams comparing D1 with a variable λ light-chain domain and a variable heavy-chain domain. The point of view is as in Fig. 2; the CDR loops of immunoglobulins and their analogues in D1 are indicated by numerals. *b*, Ribbon-diagram comparisons of D2 with a constant domain and with domain D2 of the bacterial chaperone protein PapD. *c*, Topology diagrams and strand nomenclature for the β sheets in CD4 and in variable and constant domains of immunoglobulins.

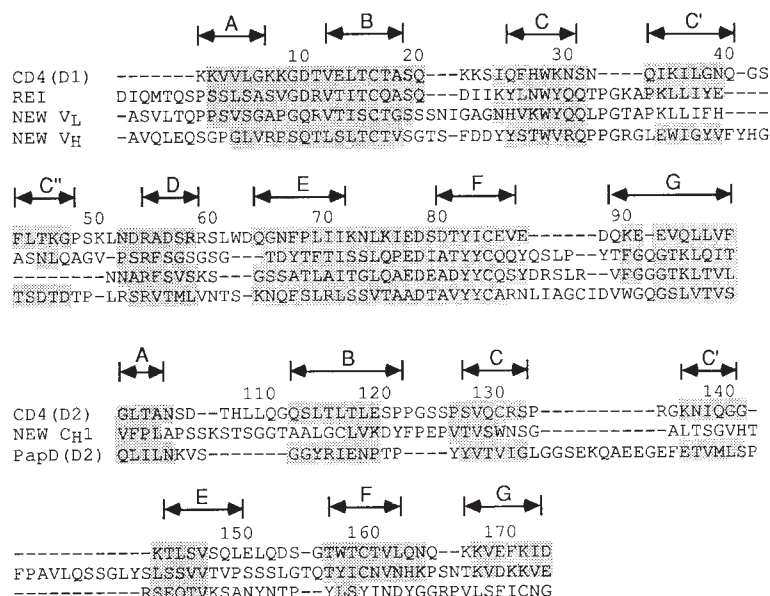
which corresponds sequentially to the D strand of normal constant domains, joins the β sheet consisting of strands C, F and G, rather than being hydrogen-bonded to strand E in the other sheet. In this respect, D2 is V-like. A final striking difference from immunoglobulins is that D2 has its disulphide bond between strands in the same sheet, that is strands C and F (Fig. 2), rather than between sheets as is usual. This feature, which is unusual but not unprecedented³⁸, had been predicted³⁴; however, strand C' was not anticipated. The hydrophobic core of immunoglobulin domains is preserved in D2 despite the unusual disulphide connection (Fig. 4). Leu 116 on strand B occupies the position of a normal cysteine partner for Cys 159 on strand F. The inward orientation of the intra-sheet disulphide bridge makes it an important member of the hydrophobic core even though it does not join the sheets.

D2 can be structurally aligned with the CH1 domain of Fab New (Fig. 4), but the resulting spatial superposition is not as close as for D1 comparisons with V domains. Only six of the seven strands superimpose at a 2.5 Å stringency for matches, and this gives 33 matches with an r.m.s. discrepancy of 1.59 Å. D2 superimposes somewhat less well with the Rei variable domain (30 matches within 2.5 Å for 1.68 Å r.m.s. discrepancy). In fact, D2 is actually more similar to the domains of PapD³⁵ than to immunoglobulin domains. As in the D2 topology, PapD domains also exhibit sheet switching (partially in D1(PapD), fully in D2(PapD)), and 39 C α atoms of D2 (CD4) superimpose on D1(PapD) with an r.m.s. discrepancy of 1.55 Å. Drawings of the aligned structures are shown in Fig. 3.

Binding site for gp120

The initial molecular event in HIV infection is the binding of gp120 on the viral envelope to CD4 on the cell surface. The high affinity of this interaction, at least for several laboratory strains of the HIV-1 virus³⁹, has permitted a detailed mapping of the binding site. More than 200 mutant CD4 recombinants have been constructed and tested for gp120 affinity (refs 8, 10–18, J. Arthos *et al.*, manuscript submitted, and unpublished results). Some of these mutant proteins have also been used to map the CD4 epitopes of a battery of monoclonal antibodies. In Table 2 we distil the mutation results to those point mutants in D1 which impair gp120 binding without causing extensive disruptions of structure as judged by antibody binding. Analyses have been reported for mutations at 80 of the 98 residues in D1, but only 19 positions among them have impact on gp120 binding without apparent global conformational change. Thirteen of these are in the span from 38 to 59. Completely buried

FIG. 4 Structural alignment of the amino-acid sequences of CD4 domains with other immunoglobulin-related domains. Shaded residues have C α positions within 2.5 Å of corresponding CD4 positions after optimal superposition of all shaded residues for a given pair of domains. (Exceptions up to 2.60 Å were allowed for residues in the middle of strands.) Each superposition relates a certain number of positions, *N*, within the specified 2.5 Å limit, and these match at a certain r.m.s. discrepancy, Δ . For the match of D1 with Rei (VL κ), *N*=72 and Δ =1.22 Å; for D1 versus New (VL λ), *N*=66 and Δ =1.12 Å; and for D1 versus New (VH), *N*=68 and Δ =1.13 Å. For the match of D2 with New (CH1), *N*=33 and Δ =1.59 Å; for D2 versus PapD(D2), *N*=32 and Δ =1.26 Å; for D2 versus PapD(D1), *N*=39 and Δ =1.55 Å; and for D2 versus Rei (VL κ), *N*=30 and Δ =1.68 Å. The alignments of D2 versus PapD(D1) and D2 versus Rei (VL κ) are not shown.



residues can be discounted as direct participants in binding because their exposure would require major conformational change, which is without precedent for immunoglobulin domains. We have calculated the degree of exposure for each residue in D1 (Table 2), and find that three of the gp120-sensitive residues, Gly 38, Ala 55 and Phe 67, are inaccessible. It seems unlikely that either Lys 29 or Gln 64 is critical for binding because there is no effect for mutations at the more exposed neighbours of Lys 29 or for single or multiple mutations at 64, except alanine. Thus only the alanine replacements at 77, 81 and 85 fall outside of the 41–59 span. The distribution of binding mutants is shown in Fig. 5*a, b*.

The importance of the 41–59 region in HIV binding is corroborated by a chimaeric CD4 in which the human sequence from 36 to 62 has been inserted into rat CD4 (A. F. Williams, personal communication). This imparts full gp120 binding activity on the normally unreactive rat CD4 molecule. It also shows that 33 additional side chains which differ in rat and human CD4, including Gln 64, are not absolutely essential for the binding⁴. But, Glu 77, Thr 81 and Glu 85 are among the residues conserved between rodents and primates. Thus, whereas the 41–59 region is positively identified as being involved in gp120 binding, participation of the 77–85 region cannot be excluded.

Antibody binding studies also help to define the gp120 binding

site (refs 10, 40, and our unpublished results), pointing especially to the CDR2-like region. Two major groups of D1-binding monoclonal antibodies have been described and their epitopes are illustrated in Fig. 5*c*. The epitope of one cross-blocking group, typified by Leu3a, maps to the CDR2-like loop and usually includes portions of the CDR1-like loop. These antibodies also block gp120 binding. The epitope of the second group, typified by monoclonal L71, includes the CDR3-like loop and sometimes a portion of the strand G. These antibodies inhibit viral infection and cell fusion, but do not efficiently block gp120 binding; some even form ternary complexes with gp120 and CD4 (A. Trunch *et al.*, manuscript submitted).

The CDR2-like region that is strongly implicated in gp120 binding is also very prominent in the CD4 structure (Figs 3 and 5). The C'–C'' hairpin loop at Gln 40–Phe 43 is almost completely exposed. Most strikingly, the hydrophobic side chain of Phe 43 juts out into the solvent as shown in Fig. 5*e, f*. The involvement of a phenyl group in binding is also suggested by peptide inhibitor studies^{41,42}. If the 77–85 span were also to be directly involved in gp120 binding, this would present a puzzle because this site is on a face nearly opposite from the CDR2-like region (the vector from the molecular centre of D1 through C α 43 makes an angle of 152° with the similar vector through C α 81), and no intervening residues have been implicated in gp120 recognition.

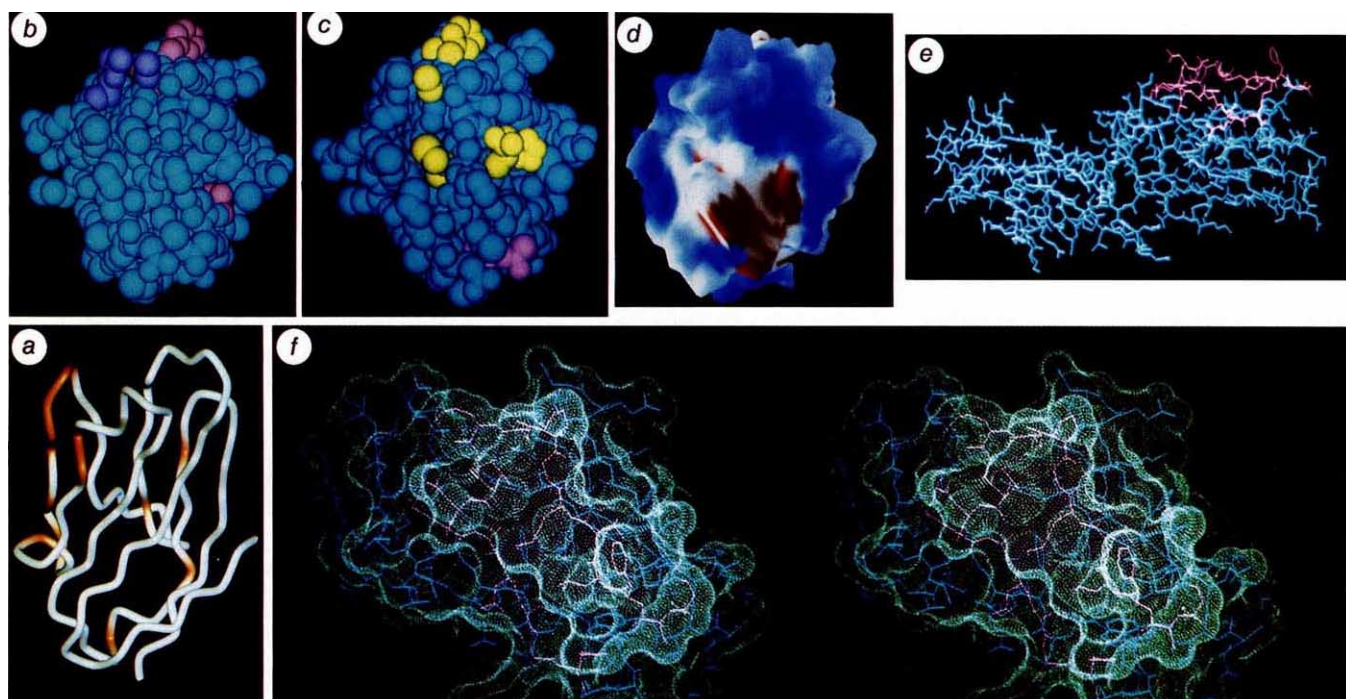


FIG. 5 Images of CD4 relating to sites of interaction with HIV. *a*, C α backbone of D1 with sites implicated by mutational and structural analysis to affect high-affinity gp120 binding (see Table 2). Non-buried residues that show reduction in binding without global disruption of structure (as judged by antibody binding) are coloured red-orange (29, 41, 42, 43, 44, 45, 47, 49, 52, 58, 59, 77, 81 and 85) on a backbone drawn by WORM (L. Andrews). The orientation is as in Figs 2 and 3. *b*, Van der Waals' surface of D1 with side chains of gp120-sensitive residues coloured. Those showing marked reductions are in pink (43, 44, 85, are visible) and those showing moderate effects are in purple (59, 42, are visible). The point of view is from above *a*, looking down at the tips of CDR-like loops. *c*, Van der Waals' surface of D1 with side chains of antibody epitopes coloured. Residues identified with epitopes of the Leu3a family and shown in purple (24, 25, 27, 42, 43, visible) and those identified with the L71 family are shown in yellow (88, 89, visible). The view is roughly as in *b*. *d*, Electrostatic potential surface computed with DELPHI³³ at neutral pH, and displayed with AAK (A. Nicholls and B. Honig)

at the levels of the solvent-accessible surface with a probe radius of 1.4 Å. Blue represents positive potential, red negative, and white neutral. The negative patch is associated with residues 85, 87 and 88. The view is approximately as in *b, e*. *e*, An all-atom representation of D1D2. Residues in the gp120-binding region from residue 41 to 59 are drawn in red. The direction of view is roughly as in Fig. 2, but the molecule has been rotated by ~90° about this view axis. This view illustrates the exposed nature of Phe 43. The actual conformation of this phenyl side chain in the crystal is at least partly determined by lattice interactions, but its highly exposed nature would also be expected to persist in other conformations accessible in solution. *f*, Stereoview of the molecular surface in the major gp120-binding region of D1. Atoms are drawn as in *e*, and are enveloped by the surface in contact with a probe sphere of 1.4 Å radius as displayed by QUANTA (Polygen). The view is taken after rotation from *e* by ~90° about the horizontal axis.

Fusion determinants

After the initial binding event, entry of virus into the cell occurs through fusion of cellular and viral envelope membranes. In addition, HIV-infected cells fuse with uninfected CD4⁺ cells to form syncytia, a process mediated by interactions between HIV envelope protein expressed on the infected cell surface and CD4 on the uninfected cell. Syncytium formation is thought to mimic viral entry and could be an important mode of viral spreading. Chimpanzee CD4, which binds to HIV but does not support syncytium formation, has Gly replacing Glu at residue 87. This replacement in human CD4 abolishes syncytia formation while preserving gp120 binding¹⁷ and, curiously, normal viral infection kinetics. Residues 88 and 89 are also implicated in this process by mutation (Table 2). These residues lie at the β -turn tip of the CDR3-like loop in D1 and are part of a patch of negative potential on the otherwise mostly positive D1D2 surface (Fig. 5d). The CDR3-like loop is spatially separated from the CDR2-like loop implicated in high-affinity binding to gp120 (C α 87 is 17 Å from C α 43). In accord, monoclonal antibodies of the L71 family block syncytium formation although they permit gp120 binding to CD4. So, it seems that the CDR3-like loop is necessary for secondary interactions between the viral envelope and CD4 before fusion. This could relate to CD4-induced release of gp120 from virus and infected cells⁴³⁻⁴⁵ and exposure of a fusogenic domain on the membrane envelope protein gp41.

Immune response interactions

The observation that CD4⁺ lymphocytes react only with class II-bearing target cells suggests that CD4 associates directly with class II MHC, and it is thought that CD4 may also have a role in signal transduction. There is evidence from cellular assays to support the association of CD4 with class II MHC molecules⁴⁶, the T-cell receptor⁴⁷, and the p56^{lck} kinase⁴⁸, as well as other CD4 molecules⁴⁹. Except for the interaction of CD4 with p56^{lck}, however, these associations are necessarily weak and thus difficult to measure. Mutagenesis *in vitro* has been used to identify the sites on CD4 that interact with class II MHC molecules. On the basis of assays of either cell adhesion^{18,50} or interleukin-2 production¹⁸, large separated expanses of the CD4 molecule, involving regions on D1, D2 and D3, have been implicated in the CD4-MHC interaction. These findings are not easy to reconcile with the structure of CD4. Some of the disruptive mutations occur in contact or bridging regions between the domains. Perhaps some of the other disruptions reflect complications from the diverse components of the cellular system rather than direct binding. For example, self-associations of CD4¹⁹ might be involved in signal transduction and possibly in class II binding.

Evolutionary implications

There is little doubt that CD4 and immunoglobulins have evolved from a common ancestor. Intron and domain boundaries coincide and, although the sequences are highly divergent, superimposable strand topologies and a common hydrophobic core are preserved. It seems likely, however, that evolution to the antibody family, with its vast repertoire of dimeric receptor units, was a later event, and that CD4 is more prototypical of immunoglobulin superfamily receptors which are often monomeric and nonpolymorphic. The absence from CD4 of J-like regions, which impart diversity in antibodies, is consistent with this. The progenitor of the diverse immunoglobulin superfamily of the present may have its vestige in the gene structure of D1, which is split by an intron (at position 47) as are genes for other immunoglobulin-like domains^{51,52}. A quasi-diad axis, perpendicular to the sheets and passing midway between strands B and E on one sheet and between strands C and F on the other, can be used to superimpose successive strands in the first exon on those in the second exon. An immunoglobulin progenitor produced by a gene duplication event⁵³ might have been

V-like, but a D2-like progenitor that would evolve to V and C domains is also a possibility.

Whatever the course of evolution, it is remarkable that molecules that function in recognition are often members of the immunoglobulin gene family. Clearly the immunoglobulin fold provides a facile framework from which a variety of loops with distinctive characteristics can be elaborated. These loops can have distinct functions, as in binding to gp120 at the CDR2-like loop and affecting fusion from the CDR3-like loop, and such modularity might have evolutionary advantage. Similarly, the quasi-symmetric nature of the immunoglobulin motif, with N and C termini at opposite ends, facilitates the concatenation of tandem, flexibly linked modules that can have distinct roles. In the case of CD4, one domain might be involved in class II binding whereas another domain might effect self-associations of the kind found in crystalline polymorphs¹⁹, and these could be essential for signal transduction. Indeed, in an alignment of D3 with D1, dimerization loops of immunoglobulin V domains would not be foreshortened in D3 as they are in D1. Such modules could evolve separately and be recombined by exon shuffling to integrate complex recognition and effector functions. □

Received 12 October; accepted 29 October 1990.

- Robey, E. & Axel, R. *Cell* **60**, 697-700 (1990).
- Maddon, P. J. *et al. Cell* **42**, 93-104 (1985).
- Williams, A. F. & Barclay, A. N. *Rev. Immun.* **6**, 381-405 (1988).
- Clark, S. J., Jefferies, W. A., Barclay, A. N., Gagnon, J. & Williams, A. F. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1649-1653 (1987).
- Maddon, P. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 9155-9159 (1987).
- Dalglish, A. G. *et al. Nature* **312**, 763-767 (1984).
- Klatzmann, D. *et al. Nature* **312**, 767-768 (1984).
- Arthos, J. *et al. Cell* **57**, 469-481 (1989).
- Capon, D. J. *et al. Nature* **337**, 525-531 (1988).
- Peterson, A. & Seed, B. *Cell* **54**, 65-72 (1988).
- Landau, N. R., Warton, M. & Littman, D. R. *Nature* **334**, 159-162 (1988).
- Clayton, L. K. *et al. Nature* **335**, 363-366 (1988).
- Mizukami, T., Fuerst, T. R., Berger, E. A. & Moss, B. *Proc. natn. Acad. Sci. U.S.A.* **85**, 9273-9277 (1988).
- Chao, B. H. *et al. J. biol. Chem.* **264**, 5182-5186 (1989).
- Brodsky, M. H., Warton, M., Myers, R. M. & Littman, D. R. *J. Immun.* **144**, 3078-3086 (1990).
- Ashkenazi, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 7150-7154 (1990).
- Camerini, D. & Seed, B. *Cell* **60**, 747-754 (1990).
- Lamarre, D. *et al. Science* **245**, 743-746 (1989).
- Kwong, P. D. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 6423-6427 (1990).
- Davis, S. J. *et al. J. molec. Biol.* **213**, 7-10 (1990).
- Richardson, N. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 6102-6106 (1988).
- Ibegbu, C. C. *et al. J. Immun.* **142**, 2250-2256 (1989).
- Healey, D. *et al. J. exp. Med.* **172**, 1233-1242 (1990).
- Chamow, S. M. *Biochemistry* **29**, 9885-9891 (1990).
- Wang, J. *et al. Nature* **348**, 411-418 (1990).
- Carr, S. *et al. J. biol. Chem.* **264**, 21286-21295 (1989).
- Xuong, N. H., Sullivan, D., Nielsen, C. & Hamlin, R. *Acta crystallogr.* **B41**, 267-269 (1985).
- Ashford, V., Dai, X., Nielsen, C., Sullivan, D. & Xuong, N. H. *Acta crystallogr.* **A46**, C-15 (1990).
- Jones, T. A. *J. appl. Crystallogr.* **11**, 268-272 (1978).
- Hendrickson, W. A. & Konnert, J. H. in *Computing in Crystallography* (eds Diamond, R., Ramaseshan, S. & Venkatesan, K.) 13.01-13.23 (Indian Institute of Science, Bangalore, 1980).
- Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458-460 (1987).
- Saul, F. A. *et al. J. biol. Chem.* **253**, 585-597 (1978).
- Gilson, M., Sharp, K. & Honig, B. *J. comp. Chem.* **9**, 327-335 (1987).
- Williams, A. F., Davis, S. J., He, Q. & Barclay, A. N. *Cold Spring Harb. Symp. quant. Biol.* **54**, 637-647 (1989).
- Holmgren, A. & Brändén, C.-I., *Nature* **342**, 248-251 (1989).
- Epp, O. *et al. Eur. J. Biochem.* **45**, 513-524 (1974).
- Bates, P. A., McGregor, M. J., Islam, S. A., Sattentau, Q. & Sternberg, M. J. E. *Protein Eng.* **3**, 13-21 (1989).
- Wilson, I. A., Skehel, J. J. & Wiley, D. L. *Nature* **289**, 366-373 (1981).
- Daar, E. S., Xi, L. L., Moudgil, T. & Ho, D. D. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6574-6578 (1990).
- Sattentau, Q. J. *et al. J. exp. Med.* **170**, 1319-1334 (1989).
- Lifson, J. D. *et al. Science* **241**, 712-716 (1988).
- Finberg, R. W. *et al. Science* **249**, 287-291 (1990).
- Kirsch, R. *et al. AIDS Res. hum. Retroviruses* **6**, 1209-1212 (1990).
- Moore, J. P., McKeating, J. A., Weiss, R. A. & Sattentau, Q. J. *Science* (in the press).
- Hart, T. K. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Doyle, C. & Strominger, J. L. *Nature* **330**, 256-259 (1987).
- Chuck, R. S., Cantor, C. R. & Tse, D. B. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5021-5025 (1990).
- Turner, J. M. *et al. Cell* **60**, 755-765 (1990).
- Veilleite, A., Bookman, M. A., Horak, E. M., Samelson, L. E. & Bolen, J. B. *Nature* **338**, 257-259 (1989).
- Clayton, L. K., Sieh, M., Pious, D. A. & Reinherz, E. L. *Nature* **339**, 548-551 (1989).
- Owens, G. C., Edelman, G. M. & Cunningham, B. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 294-298 (1987).
- Lemke, G., Lamar, E. & Patterson, J. *Neuron* **1**, 73-83 (1988).

53. Bourgois, A. *Immunochemistry* **12**, 873-876 (1975).
 54. Hendrickson, W. A. *Trans. Am. crystallogr. Assoc.* **21**, 11-21 (1985).
 55. Pähler, A., Smith, J. L. & Hendrickson, W. A. *Acta crystallogr.* **A46**, 537-540 (1990).
 56. Karle, J. *Int. J. quant. Chem. Symp.* **7**, 357-367 (1980).
 57. Hendrickson, W. A. *Acta crystallogr.* **B27**, 1472-1473 (1971).
 58. Hendrickson, W. A., Klippenstein, G. L. & Ward, K. B. *Proc. natn. Acad. Sci. U.S.A.* **72**, 2160-2164 (1975).
 59. Sheriff, S., Hendrickson, W. A., Stenkamp, R. E., Sieker, L. C. & Jensen, L. H. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1104-1107 (1985).

ACKNOWLEDGEMENTS. We thank M. Skinner, D. Sullivan, V. Ashford and C. Neilsen for help in data collection; colleagues at SmithKline Beecham for help in protein preparation and purification; C. Ogata for helpful discussions; Z. Otwinowski for his phasing program; C. Brändén and M. Amzel for supplying PapD and Fab New coordinates, respectively; A. Williams, Q. Sattentau, P. Bjorkman and D. Capon for communicating results before publication, and R. Yarmolinski for her drawing. This study was supported in part by the National Institutes of Health (W.A.H.) and by the Markey Foundation (N.-h.X.); the area detector facility at UCSD is an NIH Research Resource (N.-h.X.). Atomic coordinates have been deposited in the Protein Data Bank with entry name 1CD4 for release one year after publication of this paper.

LETTERS TO NATURE

Entropy production as the selection rule between different growth morphologies

Adrian Hill

Physiological Laboratory, Downing Street, Cambridge CB2 3EG, UK

CRYSTALLIZATION of a solid phase from a melt or solution is a special case of pattern formation in which the dissipation of energy across the free energy gradient between the two phases can give rise to various growth morphologies in the steady state¹. Experimental studies of crystallization from undercooled solutions², electrolytic deposition^{3,4} and the formation of fluid patterns in a Hele-Shaw cell⁵ have revealed faceted, dendritic, 'dense-branching' and fractal morphologies⁶. For a system with fixed anisotropy and interfacial tension, changes in the driving force for the transition (such as the degree of undercooling) can cause changes in growth morphology which are usually accompanied by changes in growth rate. The selection rule that determines these morphologies remains unclear, although a recent suggestion^{5,6} is that it is based on the growth velocity. Here I propose that selection is governed by the rate of entropy production per unit area of the different growth patterns. This principle allows accurate prediction of the morphology transition observed for the crystallization of NH₄Cl (ref. 2). I suggest that it may reflect a more general thermodynamic principle underlying a wide range of natural processes.

Consider the formation of a generalized 'crystal' in which the driving force X (ignoring surface terms) is proportional to either supersaturation, pressure or other fields. The form of the crystal involves many surface orientations which have a different surface free energies. Denoting this surface free energy f by

$$f = u - Ts \quad (1)$$

where u and s are the surface energy and entropy of the crystal, the velocity V representing the rate of crystallization is expressed as a linear function

$$V = L(X - \theta) \quad (2)$$

where X is the difference between the free energies of the dissolved solute and the bulk solid and θ is the free energy required to increase the area, so that $X - \theta$ is the total driving force, and L is the rate coefficient reflecting the growth rate of a particular morphology and its rate of entropy production. The dissipation is given by

$$\phi = dS/dt = V(X - \theta) = L(X - \theta)^2 \quad (3)$$

where θ is a linear function of f and S is the entropy created in the crystallization. A model of the process of growth that relates velocity to driving force is a model of the rate coefficient and a link between the kinetics and the entropy production. Ben-Jacob and Garik⁶ suggest that there must be a conjugate variable to the average rate of crystal growth. This conjugate variable is the force $(X - \theta)$. Equations (2) and (3) assume linear phenomenological relations but these are merely con-

venient and not essential: they almost certainly do not hold far from equilibrium or when chemical reactions occur.

The dissipation functions of two morphologies formed by adjacent steady states will in general have crossover points (Fig. 1). It follows directly from equation (3) that the crossover point for adjacent morphologies occurs at the driving force

$$X_c = \frac{(L_1\theta_1 - L_2\theta_2) + (\theta_1 - \theta_2)\sqrt{(L_1L_2)}}{(L_1 - L_2)} \quad (4)$$

There are entropic crossovers because equation (3) is a parabola in the $\phi - X$ plane: the section between $X = 0$ and $X = \theta$ represents a disaggregation of the crystal structure. This does not normally occur in experiments unless crystals are initially present, in which case one morphology disappears as the other is formed. As ϕ is always positive, I consider for the moment only the parts of the entropy curves with positive gradients because these represent crystallization, that is, the forward reaction driven by X . Reference to Fig. 1 shows that if a crossover point is to occur in this domain when $\theta_1 < \theta_2$ then $L_1 < L_2$. At higher driving forces new crossovers occur if there are steady states that can exist by creating other morphologies with higher values of L . When the driving force is reduced from a high value, the principle of selection for the steady state with highest value of ϕ predicts that at the crossover points (c_1 and c_2 in Fig. 1) transitions occur to different steady states producing different crystal morphologies. In the linear regime these transitions cannot occur without discontinuities in the rate of crystallization because, as follows from equations (2) and (4), the only point free from such discontinuity is the trivial one

$$(\theta_1 - \theta_2)\sqrt{(L_1L_2)} = 0 \quad (5)$$

in which the rates are zero. Small differences in entropy production rates at a crossover, however, could lead to small changes in slope together with small discontinuities which may be masked by variation in the data. Discontinuities are indeed a striking feature of morphology changes in various experimental

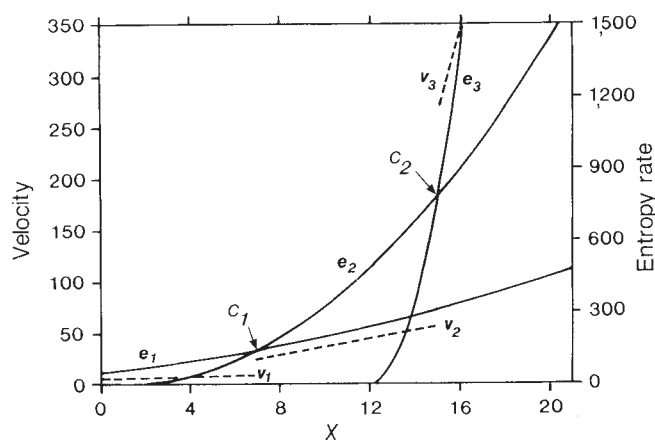


FIG. 1. The crossing points c_1 and c_2 of the entropic curves e_1 - e_3 (solid lines) of three morphologies. The associated velocities v_1 - v_3 (dashed lines) are also shown. The constants for the three entropic curves are $L_1=0.5$, $\theta_1=-10$; $L_2=4$, $\theta_2=1$; $L_3=87$, $\theta_3=12$.

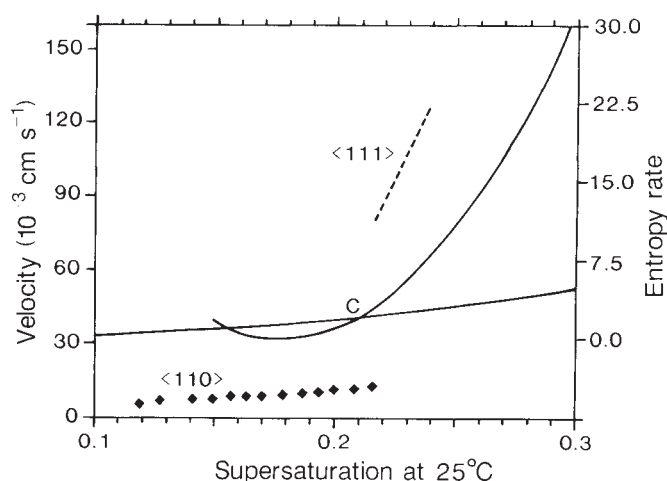


FIG. 2 Calculated entropic crossover in the change of crystal morphology between $\langle 110 \rangle$ and $\langle 111 \rangle$ dendritic forms of NH_4Cl (diamonds and dashed line). Data taken from Chan *et al.*², Figs 2 and 3. Linear regression lines were fitted to the data and the values of L and θ calculated according to equation (2): $\langle 110 \rangle$, $L=67.79$, $\theta=0.031$; $\langle 111 \rangle$, $L=2040.3$, $\theta=0.177$. The entropy curves (solid lines) were calculated in arbitrary entropy units. Entropies may be calculated for the crystal surface as $\text{J cm}^{-2} \text{ s}^{-1}$ by multiplying the velocity by $\sigma \rho k$ where σ is the specific entropy of fusion, ρ the density of NH_4Cl and k is a dimensionless geometric factor for the dendrite that relates linear growth to accretion area; all three were assumed equal for the different dendritic forms. The crossover c appears at a supersaturation of 0.21 which differs by less than 3% from the observed discontinuity at ~ 0.216 . The entropy production below the minimum of the $\langle 111 \rangle$ entropy curve gives rise to another crossover representing the entropy rate of (potential) crystal disaggregation.

systems and it does seem that larger discontinuities are accompanied by larger changes in rate^{2,3,5}.

The calculation of the number of morphologies peculiar to a dissipative system may be so complex as to render it difficult, if not impossible, to predict transitions with any precision from first principles. If one assumes that the state of highest entropy production is selected, equations (3) and (4) can be used to calculate X_c if the dependence of velocity on the driving force X can be calculated or measured for adjacent morphologies. Although there are many examples of changes in rate at the changeover between morphologies, the experimental data for these transitions is rarely precise enough for calculation.

One example where this is possible is the change from $\langle 110 \rangle$ to $\langle 111 \rangle$ orientation in dendritic forms of NH_4Cl crystallizing from solution as a function of supersaturation^{2,7}. The considerable change in the rate of growth of the main axis is shown as a function of supersaturation in Fig. 2. Fitting these rates to equation (2) by linear regression gives values for L and θ before and after the morphology transition from which the entropy production curves and the crossover can be calculated using equations (3) and (4). Although this assumes that the data is linear (which is the case here), a suitable nonlinear fit could be applied otherwise. Allowances must be made for possible differences between the two morphologies, but neglecting these predicts an entropic crossover virtually at the morphology transition within experimental error (Fig. 2). This transition point can be derived from the two velocity curves solely by applying the maximum entropy principle in a relatively straightforward manner, and is not related to the point of intersection of the velocities, as it would be if growth rates were determining the selection.

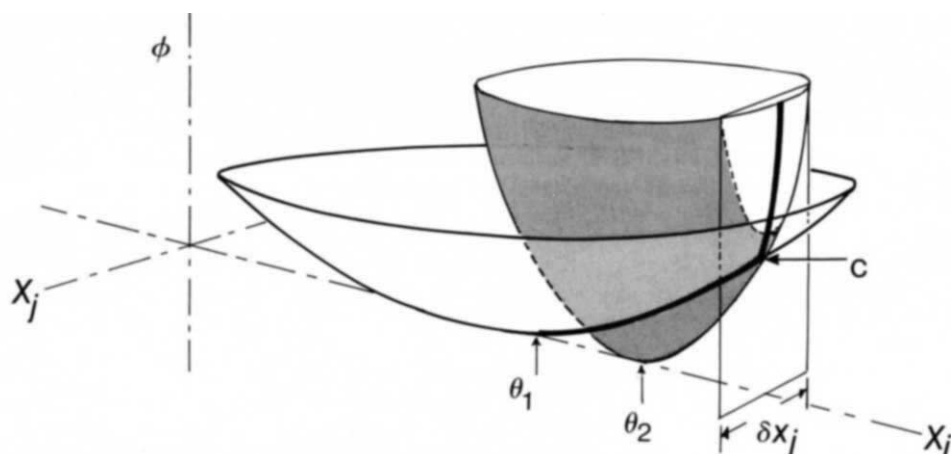
Furthermore, the velocity of the $\langle 110 \rangle$ form at the transition can be predicted to rise in a quasilinear manner with relative

supersaturation between 10 and 30 °C, as found experimentally⁷. There are changes in morphology between the $\langle 100 \rangle$ and $\langle 110 \rangle$ forms that occur at lower values of relative supersaturation but these are not distinct enough in rate to enable an analysis to be made. They do, however, exemplify the principle that two morphologies with similar values of L and θ can lead to entropic crossovers given by equation (4) with barely detectable discontinuity.

Quasicrystal formation can be described on the principle of maximum entropy selection quite differently than is usual. A problem has been to understand why quasicrystals should form at high driving forces in rapidly cooled melts. The recent theoretical demonstration that quasicrystals have high entropy^{8,9} is important because it shows that quasicrystalline morphologies are stable; it does not yet explain why they should form in preference to others far from equilibrium. I suggest that the quasicrystalline morphology is a state of higher L and θ which is therefore only manifest at high driving forces above an entropic crossover. Of all the possible crystal forms, the form selected at any driving force is not the most stable, but that which in its formation releases entropy from the free energy gradient at the highest rate. Thus neither energy nor entropy calculations of crystal forms (free energy functions θ) indicate what will form under particular conditions: the problem is one of dynamic selection based on ϕ . Possible forms for which the value of L is not high enough will never be selected. If this is correct, the problem of predicting the products of crystallization^{10,11} must involve calculations of rate constants for formation.

The principle of selection outlined here implies that transitions are always into states of greater entropy production. This does not conflict with the generally accepted principle that a

FIG. 3 Entropy production during a morphology transition. The two paraboloids represent entropy production surfaces in the space of the driving force X_i and other unconstrained forces X_j according to equation (3). At any value of the driving force the entropy production is a minimum with respect to perturbations in X_j (shown here as cutting the steeper paraboloid) but as a function of X_i always follows the path of maximum ϕ (heavy line). The curve of this path is continuous and monotonic in the $\phi - X_i$ plane with a change in slope at the crossover point c .



system in a steady state driven by a fixed force X_i is characterized by minimum entropy production¹². Any other coupled flows conjugate to forces X_j decay to zero and perturbation of any of these unconstrained forces only increases the entropy production because $\partial\phi/\partial X_j = 0$. The steady states between entropic crossovers will satisfy both conditions as can be seen from Fig. 3 for two adjacent morphologies with different L and θ values. The entropy production lines of Fig. 1 are paths of minimum ϕ within the domain of the prevailing steady state and $\partial\phi/\partial X_j = 0$ at every point on the path.

Abrupt transformations between steady states have usually been treated as a stability problem in which perturbations far enough away from equilibrium become magnified and lead to new states with order maintained by the throughput of energy. At crossover points there are kinetic instabilities that appear as changes in organization and morphology, but there is no problem of thermodynamic 'stability' with respect to the driving force X_i because the entropy production path apparent in Fig. 3 is both continuous and monotonic. $\square$

Received 27 June; accepted 8 October 1990.

1. Nicolis, G. & Prigogine, I. *Self-organization in Nonequilibrium Systems* (Wiley-Interscience, New York, 1977).
2. Chan, S. K., Reimer, H. H. & Kahlweit, M. *J. Cryst. Growth* **32**, 303–315 (1976).
3. Sawada, Y., Dougherty, A. & Gollub, J. P. *Phys. Rev. Lett.* **56**, 1260–1263 (1986).
4. Grier, D., Ben-Jacob, E., Clarke, R. & Sander, L. M. *Phys. Rev. Lett.* **56**, 1264–1267 (1986).
5. Ben-Jacob, E., Garik, P., Mueller, T. & Grier, D. *Phys. Rev. A* **38**, 1370–1380 (1988).
6. Ben-Jacob, E. & Garik, P. *Nature* **343**, 523–530 (1990).
7. Kahlweit, M. *J. Cryst. Growth* **6**, 125–129 (1970).
8. Widom, M., Deng, D. P. & Henley, C. L. *Phys. Rev. Lett.* **63**, 310–313 (1989).
9. Strandburg, K. J., Tang, L.-H. & Jaric, M. V. *Phys. Rev. Lett.* **63**, 314–317 (1989).
10. Maddox, J. *Nature* **335**, 201 (1988).
11. Hawthorne, F. C. *Nature* **345**, 297 (1990).
12. Glansdorff, P. & Prigogine, I. *Thermodynamics of Structure, Stability and Fluctuations* (Wiley-Interscience, New York, 1971).

ACKNOWLEDGEMENTS. I thank D. Hughes for drawing the figures, and B. and Y. Sachar-Hill for comments on the manuscript.

Control of product selectivity in the partial oxidation of methane

Justin S. J. Hargreaves*, Graham J. Hutchings & Richard W. Joyner

Leverhulme Centre for Innovative Catalysis, Department of Chemistry, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK

WITH the continuing need to identify new ways of synthesizing transportation fuels and hydrocarbon-based chemicals that do not rely on petroleum products, considerable effort is being devoted to the catalytic partial oxidation of natural gas. A number of oxide and other catalysts have been identified that facilitate oxidative coupling of methane to give ethane and ethene as the major products¹. But although ethene is valuable as a monomer for polymerization, C_2 hydrocarbons are of only limited usefulness. C_1 oxygenates, particularly methanol and formaldehyde, are key petrochemical intermediates, but efforts to produce them catalytically by direct methane oxidation have met with little success. Here we show that the selectivity of methane oxidation on magnesium oxide can be switched from C_2 hydrocarbons to high yields of formaldehyde simply by changing the reaction conditions, without modification of the catalyst. As well as providing insight into the mechanism for selectivity, which depends on the relative concentrations of methyl radicals and molecular oxygen, our results provide a new route to the catalytic formation of formaldehyde.

The catalytic partial oxidation of methane has been extensively studied¹. Greatest interest from the mechanistic stand-

point has been in oxidative coupling to give ethane and ethene over magnesium oxide and lithium-promoted magnesium oxide catalysts where the initial step is thought to be generation of methyl radicals at the catalyst surface^{2,3}. The reaction is normally performed at temperatures above ~ 950 K and other products observed are hydrogen, water, carbon oxides and traces of higher hydrocarbons. With most catalysts only very small amounts of methanol or formaldehyde are formed. Very high selectivities for formaldehyde ($>85\%$) have been reported at low methane conversions for some mixed-oxide catalysts, but this has only been achieved at relatively low reaction temperatures, <850 K, where coupling is not normally observed^{4–6}. There has been only a limited attempt to rationalize the experimental observations concerning product selectivities; in particular, the conditions under which the selectivities for oxygenates are high are poorly understood.

We prepared the magnesium oxide catalyst by thermal decomposition of magnesium hydroxycarbonate as previously described⁷. To reduce the surface area sufficiently to allow a wide range of oxygen conversions to be studied at convenient flow rates, the material was calcined at 1,370 K, the resulting surface area being 3 ± 1 m² g⁻¹. We used a silica microreactor (13-mm outer diameter) containing either 2.9 or 0.5 ml of catalyst. The composition of the reaction mixture was methane:oxygen:diluent in the ratio 6:1:6 at a pressure of 1 bar. Liquid products were collected downstream from the reactor and the gaseous products were analysed using gas chromatography with a thermal-conductivity detector. We normally used helium as the diluent, but argon was used when selectivities to hydrogen were to be determined. The selectivity for carbon-containing species was independent of the diluent. Liquid products were analysed by gas chromatography/mass spectrometry.

We studied partial oxidation of methane at a constant reaction temperature, 1,023 K, because selectivity is normally more sensitive to oxygen conversion than to temperature. The catalyst volume and the flow rate of the reactants were varied to give total gas hourly space velocities (ml total gas flow (at NTP) per ml catalyst per h) of $\sim 1,000$ h⁻¹ to 48,000 h⁻¹, enabling a wide range of oxygen conversions to be investigated. Figure 1 shows the variation in product selectivities with oxygen conversion; parallel trends are of course observed with methane conversion, which fell from $\sim 20\%$ to $<1\%$ as the flow rate was increased. In previous studies high oxygen conversions were used almost exclusively¹ and the selectivities reported were similar to those observed here at 90% conversion, with negligible yields of methanol or formaldehyde. As the oxygen conversion falls from 90% to 10%, we observed monotonic variations in the selectivity for most products, with the selectivity increasing for carbon monoxide, remaining constant for ethane, and decreasing for carbon dioxide and ethene. The selectivity for hydrogen does not change much. Below 10% oxygen conversion, however, we found dramatic and significant changes in selectivity—ethane disappears, hydrogen increases slightly and carbon monoxide decreases, whereas carbon dioxide remains more or less constant. Most significantly formaldehyde is the major product, indeed it seems that there is almost a direct switch from carbon monoxide and ethane to formaldehyde. The changes in selectivity with oxygen conversion are reversible and depend only on the flow rate.

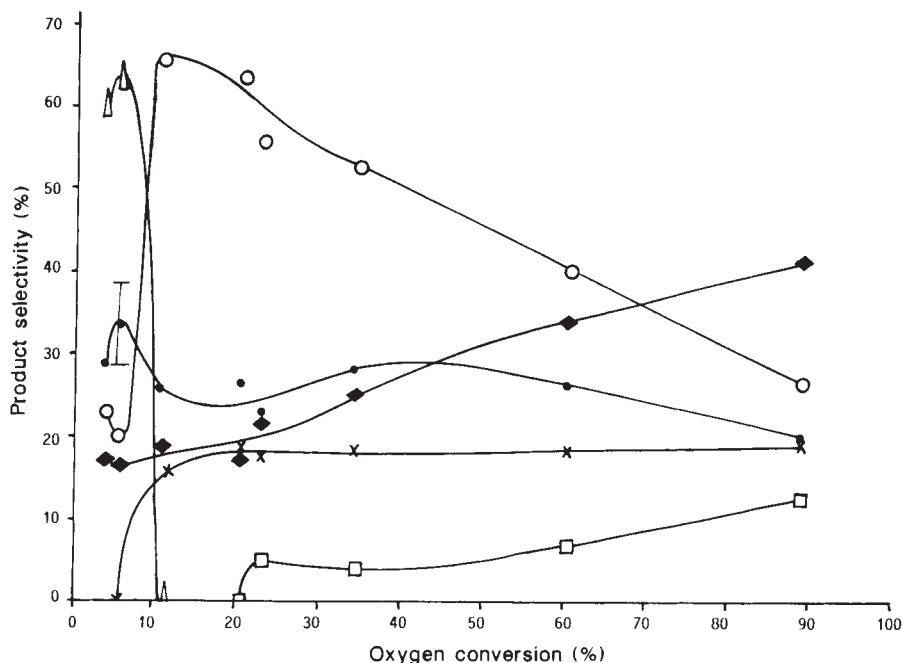
These results can be understood by considering the possible reactions of the methyl radical formed from methane by reaction with the catalyst surface³. In particular we consider the competition between coupling and oxidation, which can occur both in the gas phase and at the catalyst surface. For the coupling reaction the kinetics will be of the form



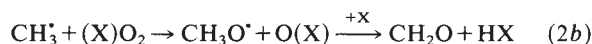
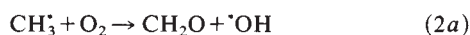
For the gas-phase oxidation, Zanthoff and Baerns⁸ have recently argued that conversion to formaldehyde will occur either directly

* Present address: Technical Chemistry Laboratory, ETH Zentrum, Zurich, CH 8092, Switzerland.

FIG. 1 Changes in selectivity in methane conversion over a magnesium oxide catalyst as a function of flow rate and oxygen conversion. X, ethane; □, ethene; ○, carbon monoxide; ◆, carbon dioxide; ●, hydrogen; △, formaldehyde. Values are accurate to $\pm 1\%$ at oxygen conversions $>60\%$, but only to $\pm 5\%$ at conversions $<10\%$. Solid lines are guides to the eye.



or, more probably⁹, via a methoxy radical



where X represents H $\cdot$ or an inert molecule. Formaldehyde may oxidize further by similar radical reactions to form carbon oxides and hydrogen or water.

The rate for all of the oxidation reactions is likely to increase as some positive power of oxygen pressure. Reactions similar to those in (2) can also be expected to occur at the surface, where ionic pathways may also contribute. The form of the surface kinetics cannot be stated with certainty, but they are also likely to show a positive dependence on oxygen pressure.

Methane is present in sufficient excess that its concentration changes little even at high oxygen conversions, so the concentration of methyl radicals generated will decrease linearly as the space velocity increases, provided the chemistry of the catalyst surface is unchanging or changing only gradually. The rate of production of ethane depends on the square of the concentration of methyl radicals. As a result, the yield of ethane (defined as the product of the extent of methane conversion, selectivity for ethane and space velocity) is expected to fall off as the flow rate is increased, as we observe. This is unusual in catalysis, where yield is often constant at high space velocities. At low methane conversion, the pressure of oxygen remains high throughout the catalyst bed, so it is not surprising that the oxidation of the methyl radical becomes dominant in comparison to the coupling reaction. We can therefore understand the shift in product selectivity from the production of C₂ hydrocarbons to formaldehyde, which is observed at $\sim 10\%$ oxygen conversions. Our experiments cannot distinguish whether the oxidation of the radical to formaldehyde occurs predominantly in the gas phase or at the surface, so significant changes in the surface chemistry with conversion cannot be ruled out. The calculations of Zanthoff and Baerns⁸ suggest that formaldehyde is not an expected product from the gas-phase oxidation of methyl radicals under conditions similar to ours, which implies that surface-catalysed oxidation is involved. If gas-phase oxidation is significant we might expect the concentration of the inert diluent to have an important role in determining how selectivity varies, because inert molecules may be necessary to carry away excess energy

released in the oxidation reactions. It is interesting and important that, over a magnesium oxide catalyst, both formaldehyde and hydrogen have some stability against oxidation.

Our results demonstrate that product selectivity for the partial oxidation of methane is controlled by the balance between methyl radical coupling and oxidation. Although we have used only undoped MgO catalysts, we consider this to be a general principle which is likely to be of importance in determining which catalysts are most useful in methane activation and in the selection of experimental conditions for the study of catalysts, particularly if oxygenated products are required in high selectivity.

Finally, we comment on the high selectivity ($\sim 25\text{--}30\%$) for hydrogen observed throughout the conversion range investigated. Three reactions are believed to occur, with the catalysed water-gas shift reaction and the thermal decomposition of ethane in the presence of water being the main pathways at high oxygen conversions. In two separate series of experiments we have found that these reactions proceed to near equilibrium conversion at the relatively low flow rates appropriate to high oxygen conversions. At higher flow rates the oxygen conversion, and hence the concentration of water, decreases, so these reactions become less significant. The main pathway to hydrogen becomes the oxidation of methyl radicals to form carbon oxides and hydrogen which again may be a homogeneous or a catalysed process. It is probable that formaldehyde is an intermediate in any gas-phase oxidation, as it is known to decompose thermally to give hydrogen and carbon oxides by a number of pathways⁸⁻¹⁰. □

Received 5 September; accepted 18 October 1990.

- Hutchings, G. J., Scurrill, M. S. & Woodhouse, J. R. *Chem. Soc. Rev.* **18**, 251-283 (1989).
- Ito, T. & Lunsford, J. H. *Nature* **314**, 721-722 (1985).
- Lunsford, J. H. *Catal. Today* **6**, 235-259 (1990).
- Mac Giolla Coda, E., Kennedy, M., McMonagle, J. B. & Hodnett, B. K. *Catal. Today* **6**, 559-566 (1990).
- Spencer, N. D. *J. Catal.* **109**, 187-197 (1988).
- Spencer, N. D. & Pereira, C. J. *J. Catal.* **116**, 399-406 (1989).
- Hargreaves, J. S. J., Hutchings, G. J. & Joyner, R. W. *Catal. Today* **6**, 481-488 (1990).
- Zanthoff, H. & Baerns, M. *Ind. Engng. Chem. Res.* **29**, 2-10 (1990).
- Walker, R. W. in *Gas Kinetics and Energy Transfer, Spec. Per. Rep. Vol. 2*, 296-330 (Chem. Soc., London, 1977).
- Geerts, J. W. M. H., Chen, Q., van Kasteren, J. M. N. & van der Wiele, K. *Catal. Today* **6**, 519-526 (1990).

ACKNOWLEDGEMENTS. We thank A. Mills for the gc/ms measurements and SERC and British Gas plc for financial support.

Bulk diffusion of metal particles on ceramic substrates

P. L. Gai*‡, B. C. Smith*§ & G. Owen†

* Department of Materials, University of Oxford, Parks Road, Oxford OX1 3PH, UK

† Film Research and Technology, Imperial Chemical Industries plc, The Heath, Runcorn WA7 4QE, UK

THE dynamical behaviour of metal particles or films deposited on ceramic surfaces is of great importance in heterogeneous catalysis, metal brazing of ceramics and microelectronics. The existing understanding of, for example, metal particle sintering or dispersion has tended to be based on static, room-temperature *ex situ* studies, which have considered only surface spreading of the metal^{1–3}. Here we present evidence from *in situ* electron microscopy for the diffusion of metal particles through the bulk of a ceramic substrate. We have studied the dynamics of small copper particles on the surface of amorphous alumina in various oxidizing and reducing atmospheres (hydrogen, oxygen and carbon monoxide). Under oxygen, particles first diffuse through the substrate with little change in shape, but at later times particles spread laterally (by surface or sub-surface diffusion) to a configuration of lower surface energy. We suggest that these diffusion processes may be important in electronics applications, which involve similar combinations of materials.

The dynamic nature of metal–ceramic interactions, a function of the environment and temperature, is crucial in determining the extent of deactivation and dispersive regeneration mechanisms in metal-particle catalysts supported on high-surface-area ceramics and in interfacial bonding in electronic devices. Metal adhesion, or wetting to the ceramic, is controlled by the substrate morphology and by the relative surface energies of the metal particle, the substrate and the metal/substrate interface. Interfacial chemical reactions are also possible.

Previous studies of metal dispersion in oxidative environments on the non-wetting or irreducible ceramics^{1–4}, such as alumina, silica (and carbon) have not mentioned any contribution of bulk diffusion through the support preceding surface spreading to form a thin film. The studies of metals on irreducible oxides, including γ -alumina^{5–7} and strongly interacting reducible oxides⁸, have been performed *ex situ* (cooling the samples to room temperature and then removing them from the reaction environment for characterization), which may not accurately represent the dynamic conditions, making it difficult to obtain insights into reaction sequences. There may also be atmospheric contamination which can cause serious errors in interpretation^{6,7}. We have therefore performed *in situ* studies in controlled gaseous environments inside a transmission electron microscope under known conditions, to probe directly the dynamic nature of metal–ceramic interactions from the same regions of sample. The relative contributions of surface and bulk diffusion have been elucidated using ESCA (electron spectroscopy for chemical analysis) and scanning electron microscopy (SEM) surface analysis techniques of the related microstructures. The *in situ* electron microscopy technique at high gas pressures and temperatures is particularly powerful for studying such systems under near operating conditions^{9–11}. Lower gas pressures may, however, be required^{12,13} to improve image resolution.

Copper–alumina is a commercially important system. It is used in catalysts for the synthesis of methanol by the hydrogenation of carbon monoxide¹⁴ at $\leq 250^\circ\text{C}$ and in water–gas shift reactions¹⁵. It is also a candidate for the fabrication of microelectronic circuits¹⁶. In the former, CO_2 and small traces of oxygen

are also present, and in the latter air oxidation is possible. We prepared the model Cu–alumina samples by electron-beam evaporation of $\sim 50 \text{ \AA}$ of Cu onto an alumina substrate $\sim 200 \text{ \AA}$ thick. High-quality non-porous amorphous alumina was prepared by anodic oxidation of aluminium. The non-porosity was confirmed by high-resolution electron microscopy. The density of amorphous alumina is similar to its crystalline equivalent. For comparison, we also prepared thin-film samples of other transition metals, noble metals and copper bimetallics, namely, Ni, Ag, Pd, Pt, Cu–Ru, Cu–Ni and Cu–Pd alloy, on non-porous amorphous silica, alumina and carbon substrates with comparable thicknesses. The materials and gases, supplied by Johnson Matthey and British Oxygen respectively, were of a spectrographically pure grade.

Figure 1a, b shows Cu–alumina treated in H_2 and O_2 , respectively, both after $\sim 1 \text{ h}$. The images were obtained directly under the typical dynamic conditions of gas pressure ($\sim 0.2 \text{ atm}$) and temperature ($\sim 200^\circ\text{C}$) and with the same tilting conditions in the microscope. Oxygen was introduced after outgassing H_2 . In H_2 , the nanometre-sized particles (for example, the one marked p in Fig. 1a) were nearly spherical, presumably their equilibrium shape. Electron diffraction confirmed that they were Cu metal. In oxygen, however, material from the particles formed diffuse rings around lighter interior regions (weak contrast marked by arrows in Fig. 1b). Further heating at the same temperature did not cause any more change and we believe that the results indicate particles in their final configuration. SEM of the top and bottom surfaces of the fully oxidized $\sim 200\text{-\AA}$ -thick sample showed particles at the bottom surface, indicating that material from the particles diffused through the substrate and subsequently spread out on the lower surface by lateral surface diffusion. Energy-dispersive X-ray analysis (EDX) showed that the lighter internal regions were devoid of Cu metal. Electron diffraction revealed CuO and Cu_2O in the outer regions of the diffuse rings. The same system exposed to CO for $\sim 1 \text{ h}$ (Fig. 1c) exhibited sub-surface diffusion of particles on subsequent exposure to oxygen, with a more complex form of spreading (Fig. 1d), which was confirmed by SEM. It is possible that the

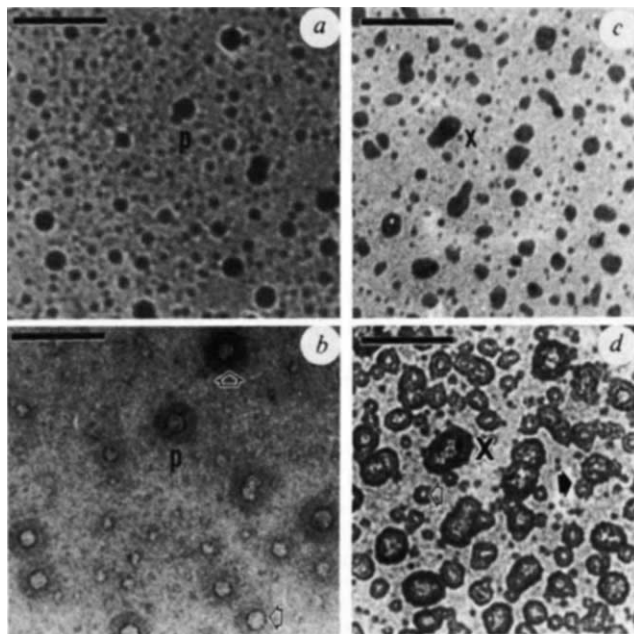


FIG. 1 Micrographs recorded under dynamic conditions at $\sim 200^\circ\text{C}$ and 0.2 atm , from the same area of Cu–alumina in a, H_2 and then b, O_2 . The diffuse regions are marked by arrows. Dynamic reaction in c, CO and then d, O_2 , with a complex form of spreading and wetting compared to that in b. Scale bar is 100 nm .

Present address: ‡ Central Research and Development Department, E. I. Du Pont de Nemours Inc., Experimental Station, Wilmington Delaware 19880-0356, USA and § Unilever plc, PO Box 760, 3000 DK Rotterdam, The Netherlands.

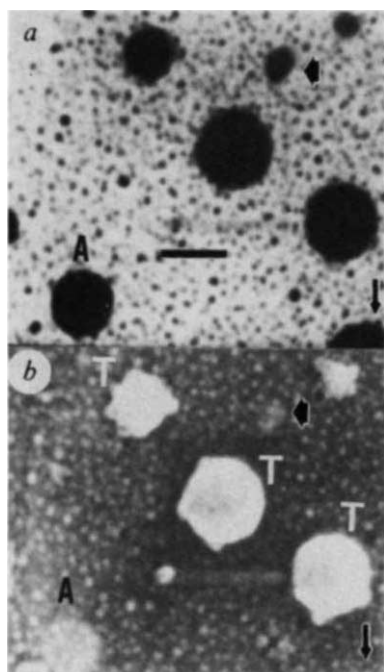


FIG. 2 *a*, STEM image (particles have the same contrast) and *b*, SEM image from the same area of a partially oxidized sample illustrating sub-surface diffusion of a copper oxide particle A (others are marked by arrows). Particles T are at the top surface. Scale bar is 20 nm.

larger CO molecules partially block the substrate surface by preferential adsorption. Conventional scanning transmission electron microscopy (STEM) and SEM images of the same region of a partially oxidized non-equilibrium sample illustrating sub-surface diffusion of particles are shown in Fig. 2*a, b*, respectively. The events refer to one stage in a continuous diffusion process before equilibrium is achieved. In the STEM image the contrast is similar for particles at the top or bottom surface of the thin (~ 200 Å thick) substrate. The SEM secondary-electron image contrast (Fig. 2*b*), however, is much stronger for particles (marked T) lying at the top surface of the thin substrate than those (for example, the one marked A) in the bulk or on the bottom surface. Particles T also have stronger contrast at the edges than at the centre, which suggests depletion of material near the centre. SEM, EDX and diffraction results support the idea² that copper oxide layers form at the outer layer of metal, but presumably subsequent diffusion of the ionic constituents of the oxide exposes successive layers of fresh metal. Although the particles seem to retain their shape as they move through the bulk, SEM at higher magnification reveals a weak contrast which we interpret as being due to spreading of particles at the surfaces. Similar diffusion processes were observed for untreated (fresh) Cu on alumina, silica and carbon in oxygen. The processes may introduce some changes in the substrate morphologies.

We believe that the different particle shapes in (Fig. 1*a, c*) are due to the differing surface energies σ of the particles. The surface energy of copper particles in hydrogen is greater than that in CO (ref. 17), because CO is adsorbed on Cu more strongly than is H₂. Under these reducing conditions, particle sintering occurred primarily by Ostwald ripening¹⁸, involving atomic migration between particles. Multiply twinned (defective) particles sintered faster.

We believe that the spreading observed under oxygen is also the result of surface-energy effects. Under equilibrium conditions

$$\sigma_{sg} - \sigma_{ms} = \sigma_{mg} \cos \theta$$

where σ are the surface energies between pairs of metal (m),

support (s) and gas (g), and θ is the contact angle between the metal particle and the ceramic. In oxidation, σ_{mg} is lowered⁴ which decreases θ , leading to the apparent increase in the radius of particles seen in Fig. 1*b, d*.

We observed both bulk and lateral diffusion in oxygen at varying temperatures for the other thin-film systems on silica, carbon and alumina, including the disordered γ -form. The results indicate that both diffusion processes must be considered in kinetic and thermodynamic treatment of oxidative dispersion. The rate of diffusion processes were found to depend on the nature of metal, substrate, thickness, gas partial pressure, temperature and time. In Cu-bimetallic systems, copper oxide particles diffused first at low temperatures.

The diffusion of copper species through the alumina substrate was confirmed by ESCA. For these experiments ~ 900 Å of Cu was electron-beam evaporated onto mica and a film of alumina ~ 200 Å thick was then deposited over the copper at a rate of ~ 100 Å min⁻¹. Oxidation studies were carried out at ~ 200 °C. Depth profiles (Fig. 3) through the samples were obtained using an argon ion beam with a constant etch rate of 2 Å min⁻¹ for ~ 300 min. Before oxidation the results showed two layers of material, with Cu beneath the alumina at the alumina/Cu interface (i_s in Fig. 3). A very high concentration of Cu²⁺ was evident after ~ 1 h of oxidation at ~ 200 °C and copper oxide was found below i_s , indicating diffusion of Cu²⁺ species into the alumina. How the bulk diffusion occurs is somewhat unclear at present but we believe that the resultant volume change (from oxidation) and reduction in σ_{mg} owing to oxidation contribute to the driving force. After ~ 16 h, Cu²⁺ was concentrated throughout the alumina near the air/alumina interface (i). Clearly, some Cu remained beneath the alumina. Using the general solution of the diffusion problem in solids¹⁹, (which is also valid for amorphous solids^{20,21}) $d \approx 2\sqrt{Dt}$ (where d is the diffusion length, D is the diffusion coefficient and t is diffusion time), we obtain an average estimate of D for diffusion of Cu species through alumina of 5×10^{-16} cm² s⁻¹. The activation energy for diffusion obtained from both the *in situ* electron microscopy and ESCA data is ~ 1.2 eV. The results indicate that oxygen promotes copper diffusion through the alumina. As the microscopy results in Fig. 1*a, c* suggest, fresh samples heated in the reducing gases

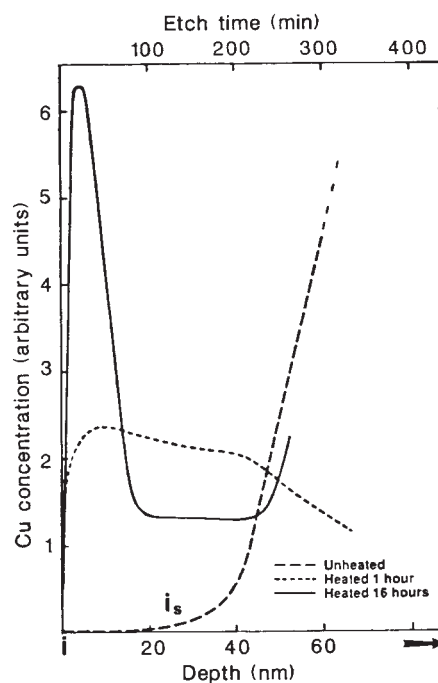


FIG. 3 ESCA depth profile indicating copper diffusion through alumina substrate following oxidation at ~ 200 °C. The sample was etched from the alumina edge (marked i) towards the alumina/Cu interface.

H₂ and CO did not show bulk diffusion of copper through the substrate. Under special conditions of intense electron-beam irradiation, however, particle penetrating has been observed in the high vacuum of an electron microscope, which is a reducing environment²².

On the basis of the newly discovered series of interactions between a reactive metal and a classically irreducible ceramic substrate at moderate temperatures in oxidative environments, we conclude that adhesion kinetics are altered by diffusion processes both into and across the substrate leading to possible loss of some metal. This has important implications in these ceramic-supported metals where only surface diffusion was previously considered. Our results show that this basic model is incomplete, and a radical revision of the interpretation of redispersion and of fundamental processes in regenerative catalysis may be required. □

Received 28 May; accepted 17 October 1990.

1. Nakayama, T., Arai, M. & Nishiyama, Y. *J. Catal.* **87**, 108–115 (1984).
2. Wang, T. & Schmidt, L. D. *J. Catal.* **70**, 187 (1981).
3. Burkhardt, J. & Schmidt, L. D. *J. Catal.* **116**, 240–251 (1989).
4. Ruckenstein, E. in *Growth of Metal Clusters* (ed. Bourdon, J.) 57–63 (Elsevier, Amsterdam, 1981).
5. Wynblatt, P. & Gjostein, N. A. *Prog. Solid. St. Chem.* **9**, 21–57 (1975).
6. Ruckenstein, E. & Lee, S. H. *J. Catal.* **86**, 457–463 (1984); **104**, 259–278 (1987).
7. Sushumna, I. et al. *J. Catal.* **94**, 239–288 (1985).
8. Tatarchuk, B. J. & Dumesic, J. A. *J. Catal.* **70**, 308–346 (1981).
9. Gai, P. L. *Phil. Mag.* **A43**, 841–855 (1981).
10. Gai, P. L. *J. Solid. St. Chem.* **49**, 25–42 (1983).
11. Boyes, E. D., Gai, P. L. & Warwick, C. *Nature* **313**, 666–668 (1985).
12. Deroune, E. G. et al. *J. Catal.* **85**, 187–196 (1984).
13. Wong, C. & McCabe, R. J. *Catal.* **119**, 47–64 (1989).
14. Madix, R. J. & Bowker, M. *Surf. Sci.* **102**, 542–553 (1982).
15. Bond, G. C. & Namijo, S. N. *J. Catal.* **118**, 507–510 (1989).
16. Gazecki, J. et al. *Surf. Sci.* **22**, 1034–1040 (1985).
17. Balooch, M. et al. *Surf. Sci.* **46**, 358–366 (1974).
18. Wanke, S. E. & Flynn, P. C. *Catal. Rev.-Sci. Engng.* **12**, 93–135 (1975).
19. Shewmon, P. G. *Diffusion in Solids* (McGraw Hill, New York, 1963).
20. Cottrell, A. H. *Introduction to Metallurgy* (Arnold, London, 1971).
21. Brebec, G. et al. *Acta metall.* **28**, 327–333 (1980).
22. Ajayan, P. M. & Marks, L. D. *Nature* **338**, 139–141 (1989).

ACKNOWLEDGEMENTS. We thank SERC and ICI for facilities and a CASE studentship (B.C.S.).

Characterization of dissolved organic matter in the Black Sea by fluorescence spectroscopy

Paula G. Coble*, Sarah A. Green†, Neil V. Blough† & Robert B. Gagosian†

* School of Oceanography, University of Washington, Seattle, Washington 98195, USA

† Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

THE natural fluorescence properties of sea water provide a means of elucidating the complex chemical composition and diverse sources of dissolved organic matter (DOM) in sea water^{1–6}. The positions of excitation and emission maxima for a wide range of natural water samples show remarkable similarity⁷. High-sensitivity fluorescence spectroscopic studies⁸ have shown recently that emission maxima for marine and coastal waters differ by 20 nm when the excitation wavelength is 313 nm. Here we present evidence from three-dimensional excitation emission matrix (EEM) spectroscopy that at least three fluorophores are present in waters of the Black Sea. Distinct changes in the relative abundance of these fluorophores are observed as a function of depth. We suggest that three-dimensional fluorescence spectroscopy can be used to distinguish between different types and sources of DOM in natural waters. These findings may have important applications in the field of remote sensing of phytoplankton pigments. For example, a better understanding of the sources of DOM components will help

in correcting^{9,10} remotely sensed data for the presence of gelbstoff (yellow-coloured DOM¹¹, which plays an important part in radiation absorption by surface waters).

We collected samples of dissolved organic matter during Leg 5 of the 1988 RV *Knorr* Black Sea Expedition (cruise 134-12, July 1988). The study site was located near the centre of the basin (43°5.0' N, 34°00.0' E) in 2,218 m of water. Dissolved oxygen was undetectable (<5 µM) below 100 m and hydrogen sulphide was present at all depths below 120 m (ref. 12). There was a constant increase in DOM fluorescence (Ex/Em = 320–390/475–530 nm) from the surface to 375 m, the maximum depth sampled^{13,14}.

DOM was concentrated from acidified (pH 3.2) sea water onto pre-cleaned octadecylsilane-bonded silica (Waters C18 Sep-Paks) at flow rates of 10–15 ml min⁻¹, eluted with methanol, concentrated by rotary evaporation at reduced pressure and redissolved in methanol^{13,15}. Absorption spectra were measured on an HP-8451 diode array spectrophotometer. Fluorescence spectra were recorded on an SLM-Aminco SPF-500C spectrofluorometer in ratio mode. A procedural blank was prepared by passing distilled water through a Sep-Pak and eluting and concentrating using the same procedures as for samples. To eliminate fluorescence contamination from the Sep-Paks and Raman scattering from the solvent (methanol), the blank EEM was subtracted from each sample EEM. Where noted, we corrected the emission spectra for instrumental configuration using correction factors provided by the manufacturer, with some modifications¹⁶.

We present EEMs before (Fig. 1) and after (Fig. 2) correction for instrumental configuration to permit comparison between our results and those of other investigators. Because uncorrected spectra were collected with the same instrument under identical conditions, they can be directly compared for differences in position and intensity of maxima even below 300 nm, where the correction routine is not valid.

Fluorescence originates from the first excited singlet state of a molecule, regardless of excitation energy. Thus for the EEM of a single simple fluorophore, the position and linewidth of the emission peak do not shift with excitation wavelength. Fluorescence intensity varies along the excitation axis, reflecting the absorption spectrum of the molecule. A cross-section through an EEM parallel to the excitation axis represents a traditional excitation scan, whereas a cross-section parallel to the emission axis is equivalent to an emission scan.

There are three distinct fluorescence maxima in the uncorrected EEMs (Fig. 1). The highest fluorescence intensities occur at excitation 260 nm and emission 440 nm (region A) for all samples except that from the 1 m depth. In the two deep-water samples, there is an emission maximum at 440 nm for excitation at 340 nm (region C). There is fluorescence in this region in the samples from 1 m and 28 m depth as well, but it is not intense enough to create a local maximum. The emission at 440 nm from excitation at two wavelengths could be interpreted as being from a single fluorophore, if the ratio between the two excitation peaks were constant. This clearly is not the case, as the ratio of fluorescence in region A to fluorescence in region C (A:C, Table 1) decreased from 1.33 in the sample from 1 m depth to 1.08 in the sample from 375 m depth. Changes in speciation due to pH variations could cause a similar effect, but the *in situ* pH varied only slightly (8.3–7.7) over the depth interval sampled¹⁷. Because all samples were acidified and the spectra measured in methanol solution, pH-induced speciation was probably not a factor. Therefore we conclude that at least two separate fluorophores are present, both of which have an emission maximum at 440 nm.

A third fluorophore is responsible for the maximum at Ex/Em = 295/345 (region B). This is the largest peak in the sample from 1 m depth and was present, but less intense, in the samples from 28 m and 375 m depth. Its fluorescence properties are similar to those of the indole ring of tryptophan¹⁸.

TABLE 1 Fluorescence maxima for natural water samples

Depth	A	B	Region C	D	E	Reference
Black Sea						
Uncorrected						
1 m	260/450 (1.2)	285/350–360 (1.4)	n.m. (0.9)			This study
28 m	260/460 (1.3)	285/350 (1.2)	n.m. (1.0)			
44 m	260/450 (1.6)	285/350 (1.1)	345/345 (1.2)			
78 m	260/455–460 (2.0)	285/360 (1.0)	345/445 (1.6)			
217 m	260/450 (2.6)	n.m. (1.0)	345/445 (2.4)			
375 m	260/445 (2.7)	n.m. (1.5)	345/445 (2.5)			
Corrected						
1 m	260/435 (0.05)	280/325–345 (0.2)	n.m. (0.03)			This study
28 m	260/435–445 (0.06)	280/345 (0.1)	n.m. (0.05)			
44 m	n.m. (0.08)	285/335 (0.1)	n.m. (0.05)			
78 m	260/445 (0.1)	n.m. (0.07)	345/445 (0.08)			
217 m	260/440 (0.2)	n.m. (0.09)	340/450 (0.18)			
375 m	260/435 (0.2)	280/335 (0.2)	340/445 (0.19)			
Suwannee River fulvic acid						
*	260/460		325/450			S. A. Green (personal communication) 24
†	230/420		300/415			
Soil fulvic acid						
‡			350/490	390/509	455/521	25

A through E indicate regions of the EEM. Actual positions of the maxima are indicated by Ex/Em wavelength pairs (in nanometres). Values in parentheses represent relative fluorescence intensities after subtraction of the blank. Fluorescence intensities in regions where no maximum was observed are indicated by 'n.m.'.

* Water, pH 8.5, 5 mg l⁻¹; † pH 4.5; ‡ 40:60 glycerol:water (w:w), pH 6.0, 150 mg l⁻¹.

FIG. 1 Uncorrected EEM spectra for concentrated DOM samples from the Black Sea at depths of 1 m, 28 m, 217 m and 375 m. A through C refer to the fluorophore designations presented in Table 1. Instrument settings were the same for each set of spectra. EEMs were generated by concatenating 40 emission spectra collected at excitation wavelengths 5 nm apart between 260 and 455 nm. Excitation and emission bandpasses were 4 nm. The EEM spectrum for a procedural blank has been subtracted from each sample spectrum. All samples were concentrated by a factor of ~67.

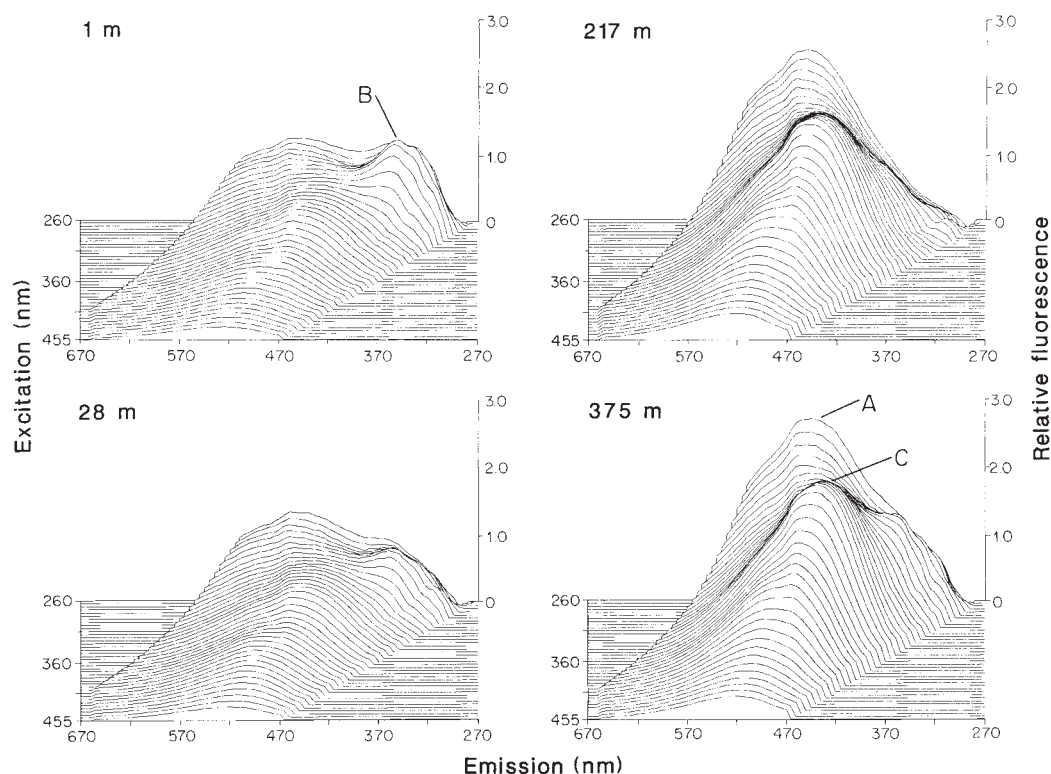
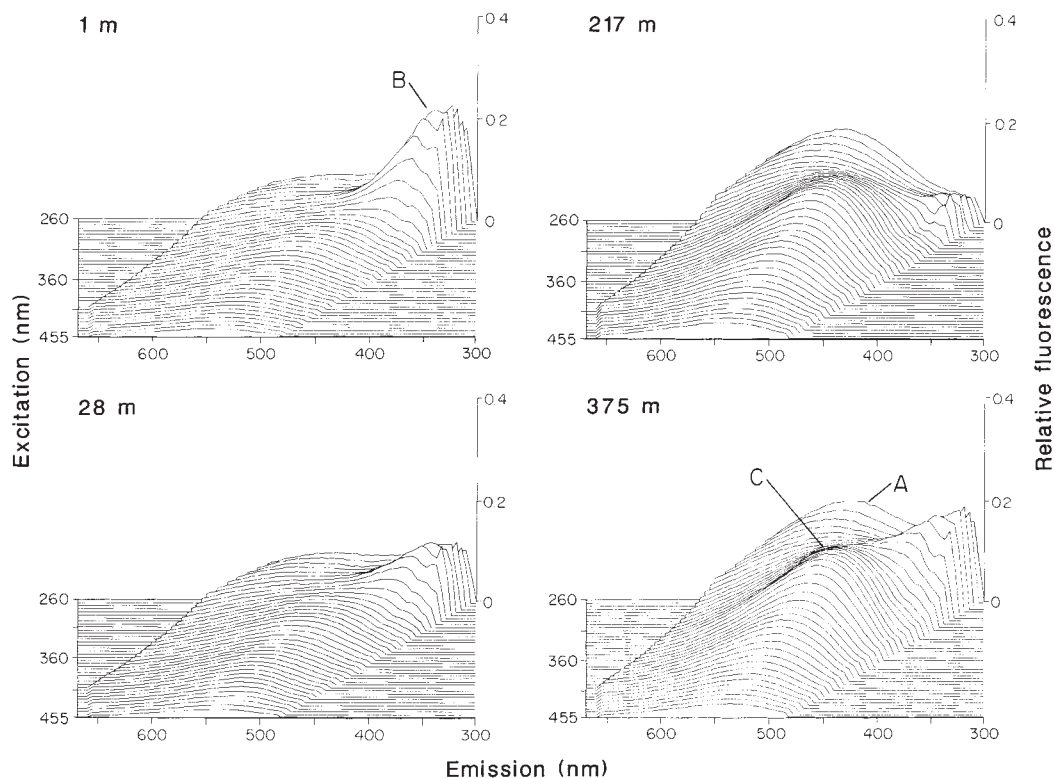


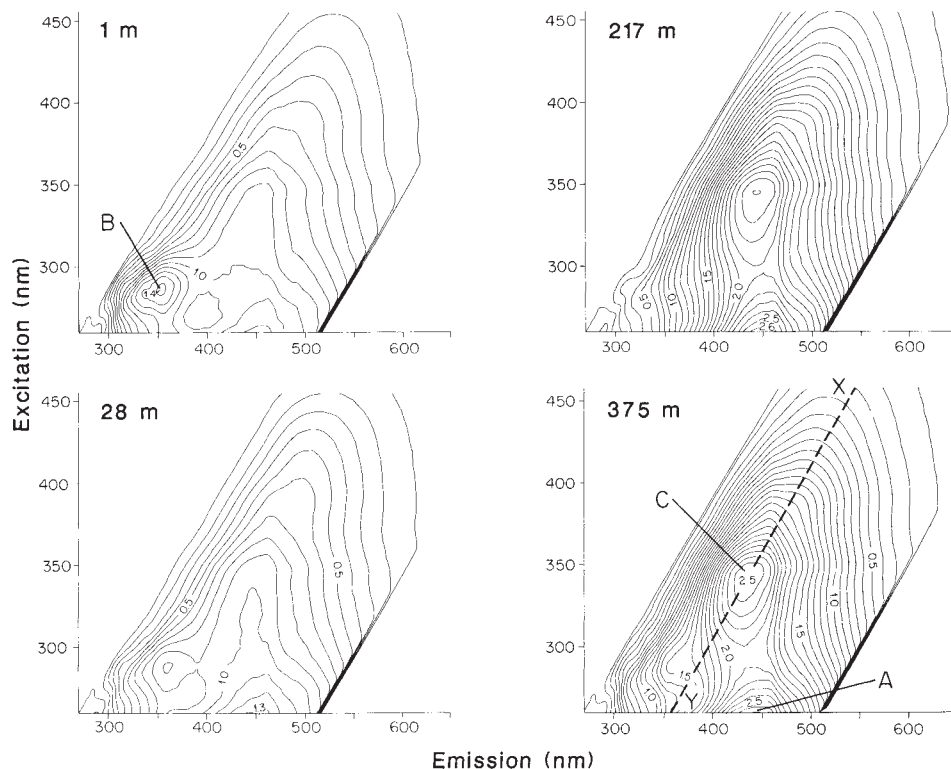
FIG. 2 Corrected EEM spectra for concentrated DOM samples from the Black Sea at depths of 1 m, 28 m, 217 m and 375 m. Spectra are the same as in Fig. 1 after correction for instrumental configuration. The procedural blank was also corrected before subtraction from samples. Data below 300 nm are lost because the correction routine is not valid below that wavelength.



The fluorescence in region C in water from all depths sampled is definitely from a mixture of fluorophores. This is indicated by the redshift in the emission maximum at increasingly longer excitation wavelengths (line X-Y; Fig. 3). Fluorescence in regions A and B may be from a single fluorophore in each case, or to the presence of multiple species having similar spectral properties.

The corrected EEMs (Fig. 2) show the same general features, but fluorescence at short-wavelength excitations is enhanced by the correction. Emission maxima in regions A and B are shifted 10–20 nm towards the blue in the corrected spectra (Table 1). Both corrected and uncorrected EEMs indicate that the ultraviolet region is important for distinguishing between samples from different depths.

FIG. 3 Contour plots of EEM spectra shown in Fig. 1 for samples from the Black Sea at depths of 1 m, 28 m, 217 m and 375 m. Contour interval is 0.1 relative fluorescence units. Line X-Y in the lower right panel represents the positions of the long-wavelength emission maxima.



The ultraviolet absorption spectra of these samples are distinctly different. The shallow samples and the sample from 375 m have a shoulder at 230 nm not seen at 217 m. We cannot relate the absorption at this wavelength to fluorescence, because our shortest excitation wavelength was 260 nm.

The apparent yield of fluorescence at 440 nm with excitation at 260 nm increases with depth. Absorbance values are close to 0.2 in all samples, but the amount of fluorescence per unit absorbance increases twofold from 1.4 in samples from 1 m to 2.7 in samples from 375 m. True quantum yields cannot be obtained because the samples contain a mixture of compounds, some of which may absorb at 260 nm but be non-fluorescent. Previous studies have shown that apparent fluorescence efficiencies are higher in coastal waters than in offshore marine waters⁸. Our observation of lower efficiencies in surface waters than in deep waters may be the result of photodegradation in the upper water column¹⁹⁻²², which decreases fluorescence intensity but may have a lesser effect on absorbance.

Excitation wavelengths shorter than 300 nm have been used in only a few previous studies, with similar results. A tryptophan-like fluorescence (Ex/Em = 287/348 nm; ref. 18) was observed for waters from a *Trichodesmium* bloom³, and a tyrosine-like fluorescence (Ex/Em = 275/303 nm; ref. 18) was observed in highly productive Antarctic waters²³. As the fluorescence signal of proteins arises from these two amino acids, fluorescence of sea water at these wavelengths could be caused by the presence of either dissolved free amino acids or dissolved proteins. The implication here is that DOM fluorescence in surface waters is associated with growth of living organisms.

Our EEMs from Black Sea waters are unlike EEMs reported from two other natural sources, both of which contained multiple fluorescing species (Table 1). A Suwannee River fulvic acid had maxima at Ex/Em = 230/420 nm and 300/415 nm (ref. 24). A soil-derived fulvic acid (Contech) had a maximum at Ex/Em = 390/509 nm, with smaller peaks at Ex/Em = 350/490 nm and 455/521 nm (ref. 25). Some of these differences may be because of differences in instrumentation, sample preparation, pH, concentration and solvents. Using our instrumentation, the EEM for Suwannee River fulvic acid (water, pH 8.5, 5 mg l⁻¹) was quite similar to the EEM for Black Sea deep water (Table 1).

These results suggest that EEM evaluation can be used to distinguish between DOM from at least three different sources: soils, rivers and surface sea water. Surface and deep sea water samples are also distinguishable, at least in the Black Sea. Preliminary results (P.G.C., unpublished data) for unconcentrated sea water samples from the Black Sea as well as from other marine and freshwater areas support this conclusion. There are some differences between the concentrated and unconcentrated samples, however, which may be caused by isolation procedures. Fluorescence in region A is less intense in unconcentrated samples, and another maximum (Ex/Em 295-300/405-410 nm) not observed in the concentrated samples is present. A fluorescence maximum at 310/410 nm has been reported previously in sea water using laser-induced fluorescence^{8,21}. The cause of the apparent difference between concentrated and unconcentrated samples needs further investigation. It could have important implications for the study of marine DOM by techniques that require isolation of the material from sea water.

Fluorescent DOM represents part of the gelbstoff signal that must be accounted for in satellite-derived estimates of phytoplankton productivity. The three types of gelbstoff in the ocean (terrestrial^{1,2}, bloom-related^{3,26-28} and condensates of non-fluorescent DOM^{2,29}) have distinct chemical compositions. EEM analysis seems to be sufficiently sensitive to distinguish between these and may be used to assess the contribution of each. □

4. Duursma, E. K. in *Optical Aspects of Oceanography* (eds Jerlov, N. G. & Steemann Nielsen, E.) 237-256 (Academic, New York, 1974).
5. Berger, P., Laane, R. W. P. M., Illhude, A. G., Ewald, M. & Courtot, P. *Oceanologica Acta* **7**, 309-314 (1984).
6. Cabaniss, S. E. & Shuman, M. S. *Mar. Chem.* **21**, 37-50 (1987).
7. Zepp, R. G. & Schlotzhauer, P. F. *Chemosphere* **10**, 479-486 (1981).
8. Donard, O. F. X., Lamotte, M., Belin, C. & Ewald, M. *Mar. Chem.* **27**, 117-136 (1989).
9. Bricaud, A., Morel, A. & Prieur, L. *Limnol. Oceanogr.* **26**, 43-53 (1981).
10. Carder, K. L., Steward, R. G., Harvey, G. R. & Ortner, P. B. *Limnol. Oceanogr.* **34**, 68-81 (1989).
11. Kalle, K. *Ann. Hydrogr.* **66**, 1-13 (1938).
12. Friederich, G. E., Codispoti, L. A. & Sakamoto, C. M. *Tech. Rept. 90-3* (Monterey Bay Aquarium Research Institute, Monterey, 1990).
13. Coble, P. G. thesis, MIT and Woods Hole Oceanogr. Inst. (1990).
14. Coble, P. G., Gagosian, R. B., Codispoti, L. A., Friederich, G. E. & Christensen, J. P. *Deep Sea Res.* (in the press).
15. Dunlap, W. C. & Susic, M. *Mar. Chem.* **17**, 185-198 (1985).
16. Green, S. G., Simpson, D. J., Zhou, G., Ho, P. S. & Blough, N. V. *J. Am. chem. Soc.* (in the press).
17. Goyet, C., Bradshaw, A. L. & Brewer, P. G. *Deep Sea Res.* (in the press).
18. Wolfbeis, O. S. in *Molecular Luminescence Spectroscopy* Part 1 (ed. Schulman, S. G.) 167-370 (Wiley-Interscience, New York, 1985).
19. Kramer, C. J. M. *Neth. J. Sea Res.* **13**, 325-329 (1979).
20. Hayase, K., Tsubota, J. & Sunada, I. *Mar. Chem.* **25**, 373-381 (1988).
21. Chen, R. F. & Bada, J. L. *Geophys. Res. Lett.* **16**, 687-690 (1989).
22. Kouassi, M. thesis, Univ. Miami (1986).
23. Ewald, M., Stabel, H. & Belin, C. *r. hebdo. Séanc. Acad. Sci. Paris* **302**, 883-886 (1986).
24. Goldberg, M. C. & Weiner, E. R. in *Humic substances in the Suwannee River, Georgia, US geol. Surv. Open-file Rep.* 87-557, 179-204 (US Dept. Interior, Washington, DC, 1989).
25. Lochmuller, C. H. & Saavedra, S. S. *Analyt. Chem.* **58**, 1978-1981 (1986).
26. Yentsch, C. S. & Reichert, C. A. *Bot. Marina* **3**, 65-74 (1961).
27. Sieburth, J. M. & Jensen, A. *J. exp. mar. Biol. Ecol.* **3**, 275-289 (1969).
28. Carlson, D. J. & Mayer, L. M. *Can. J. Fish. Aquat. Sci.* **40**, 1258-1263 (1983).
29. Hedges, J. I. *Geochim. cosmochim. Acta* **42**, 69-76 (1978).

ACKNOWLEDGEMENTS. We thank the captain, crew and scientists of RV *Knorr* cruise 134-12 and the organizers of the 1988 Joint US-Turkey Black Sea Expedition. We also thank J. Hedges for a review and L. Sylwester for assistance with preparation of the figures. This work was supported by the NSF, the Environmental Protection Agency and the Office of Naval Research. Woods Hole Oceanographic Institution Contribution 7444.

Intrusive and extrusive growth of the Mount St Helens lava dome

Jonathan H. Fink, Michael C. Malin
& Steven W. Anderson

Department of Geology, Arizona State University, Tempe,
Arizona 85287-1404, USA

LAVA domes grow by a combination of endogenous (intrusive) and exogenous (extrusive) addition of magma, either as single continuous events^{1,2} or through a series of relatively small pulses which together build a larger construct³. Evaluating the hazards that attend growth of a dome requires an assessment of the processes controlling whether lava is added to its interior or surface, because many of these dangers depend on how volatiles are admitted to or released from hot magma. If the volatile pressure inside an inflating dome becomes high enough, sudden exposure provided by flow-front slumping can lead to explosive decompression and the generation of pyroclastic flows^{4,5}. On the other hand, lava erupted on the oversteepened surface of a dome may collapse, fragment and mix with snow or ice to form pyroclastic flows, surges and mudflows^{6,7}. Here we provide a quantitative assessment of the partitioning of magma into endogenous intrusions and exogenous lobes, using high-resolution, digital topographic maps of the Mount St Helens dome derived from aerial photographs taken periodically between 1980 and 1986. Endogenous, exogenous and total volume production rates follow distinct trends which provide important clues about the nature of eruption mechanisms. Calculating endogenous and exogenous components for active domes like those recently formed at Redoubt Volcano may help quantify magma-chamber processes and provide another tool for understanding the re-equilibration of shallow magmatic systems.

The Mount St Helens dacite dome, which was active from October 1980 to October 1986, provided an excellent opportunity to evaluate the conditions necessary for these different styles of growth. Dome emplacement occurred during more than 15 eruptive pulses, or episodes, lasting from a few days to more than a year and separated by repose periods ranging from about a

Received 29 June; accepted 12 October 1990.

1. Kalle, K. *Dt. hydrogr. Z.* **2**, 117-124 (1949).
2. Kalle, K. *Oceanogr. mar. Biol. A. Rev.* **4**, 91-104 (1966).
3. Traganza, E. D. *Bull. mar. Sci.* **19**, 897-904 (1969).

month to a year. Each episode began with a period of endogenous distension, followed by extrusion of an exogenous lobe. After each eruptive event, the United States Geological Survey (USGS) acquired high-resolution aerial photographs, subsequently used to create twenty detailed topographic maps and a corresponding set of geological maps⁸. Initial evaluation of the topographic maps^{8,9} indicated regular large-scale patterns, including linear relationships between dome volume and time.

To address quantitatively questions of magma budget and the partitioning of lava between the interior and surface of the dome, we used image-processing algorithms to manipulate digital forms of these maps. The maps were digitized at three different sites (USGS field offices in Flagstaff (Arizona) and Vancouver (Washington), and the Department of Geology, Arizona State University) using slightly different techniques, so considerable calculation was required to bring them to a common format on the same base map. The resulting maps constitute a very-high-resolution set of repeatedly gathered topographic data.

The maps have 2-m contour intervals, and better than 2-m and 0.5-m vertical and horizontal resolution, respectively. We determined the elevation change associated with each episode by calculating the height difference between the topographic data sets made before and after each event. These height differences were contoured and superimposed on shaded relief versions of the pre-episode map to examine the specific portions of the dome that changed elevation during the eruptive event. Independent geodetic-point measurements (D. Swanson, personal communication) indicate that most movement involved horizontal displacements rather than vertical movement.

Distinguishing exogenous and endogenous components of a given extrusive episode required further manipulation of the pre- and post-episode topographic data sets. First, we calculated the total net volume change between the two maps. Then the boundary of the extruded lobe was determined by transferring to the shaded-relief version of the post-episode map the outline of the lobe from field observations and inspection of aerial photography. Elevation differences within this area were then integrated to calculate the exogenous volume. Finally, this exogenous component was subtracted from the total net volume change to give the endogenous volume change. This approach considers the relatively small volumes of new talus to be part of the endogenous component.

A principal source of error in this method arises from the inability to distinguish endogenous from exogenous growth within the outlines of the extruded lobe. If endogenous inflation occurs in the area that is later covered by a new lobe, these two phases of uplift are combined in the difference image. Similarly, this technique cannot account for a depression that is created and later filled. We evaluated these uncertainties by calculating the thickness that an exogenous lobe would have on a given slope if it were to possess a Bingham rheology¹⁰ with a constant yield strength. The discrepancy between calculated lobe thickness and measured height differences constitutes the endogenous inflation or other unmodelled topographic relationships within the lobe boundaries. Such calculations for Mount St Helens indicate that little endogenous inflation occurred in that portion of the dome about to receive an exogenous lobe. Other sources of error (for example, areas that initially inflate but then deflate before they are photographed) do not seem to be reflected in the overall growth behaviour of the dome.

In Fig. 1, cumulative exogenous, endogenous and total volumes are shown as functions of time since 18 October 1980. The total-volume curve, essentially equivalent to that presented by Swanson and Holcomb⁹, indicates that the long-term supply rate of magma to the dome decreased in a regular manner. Swanson and Holcomb suggested that these data reflect a steady-state magma-chamber process, such as crystallization-induced vesiculation, and pointed out that the trend allowed volume predictability in three distinct growth intervals⁹. They further

proposed that the progressive decrease in slope between each interval may have been related to significant events in the history of the dome, such as a change in lava chemistry, injection of new volatile-rich magma into the chamber, or generation of a persistent molten core in the dome.

By distinguishing endogenous from exogenous growth, we can more closely examine the long-term extrusion process. The cumulative exogenous volume reflects a decreasing production rate that has a functional form similar to that of the total volume. This trend may be described in terms of three linear segments with decreasing slopes (supply rates of 5.1×10^4 , 3.0×10^4 and $8.3 \times 10^3 \text{ m}^3 \text{ d}^{-1}$) and breaks occurring in March 1982 and February 1984, as proposed by Swanson and Holcomb⁹; or logarithmically according to a relationship between volume and the square root of time. In contrast, endogenous dome growth is remarkably predictable: the quantity of magma added to the dome interior is linearly proportional to the time since the preceding eruptive phase, and the long-term endogenous production rate is essentially constant at $2 \times 10^4 \text{ m}^3 \text{ d}^{-1}$ ($0.007 \text{ km}^3 \text{ yr}^{-1}$).

These distinct trends indicate that endogenous and exogenous growth occur by different processes. Endogenous growth occurs at a constant rate which seems relatively insensitive to changes in dome geometry, vigour or explosiveness of eruptive events, or magma-chamber processes; it requires a controlling factor deep in the dome interior or conduit. The declining exogenous production rate mimics that for total volume, possibly reflecting conditions in the source region which control the overall generation of magma.

We envision three stages in the growth of the dome which in turn control the partitioning of magma into endogenous and exogenous components: (1) crystallization in the chamber periodically generates volatile-enriched batches of magma with sufficient buoyancy to overcome lithostatic forces and rise towards the surface¹¹⁻¹⁴; (2) each of these eruptive episodes begins with an endogenous pulse that distends the dome interior, stretching the dome's deformable crust; (3) tension in this crustal layer creates micro-cracks that eventually cause it to fail at or near its geometric apex^{15,16}, allowing that portion of the buoyant magma batch remaining in the conduit to be extruded as an exogenous lobe.

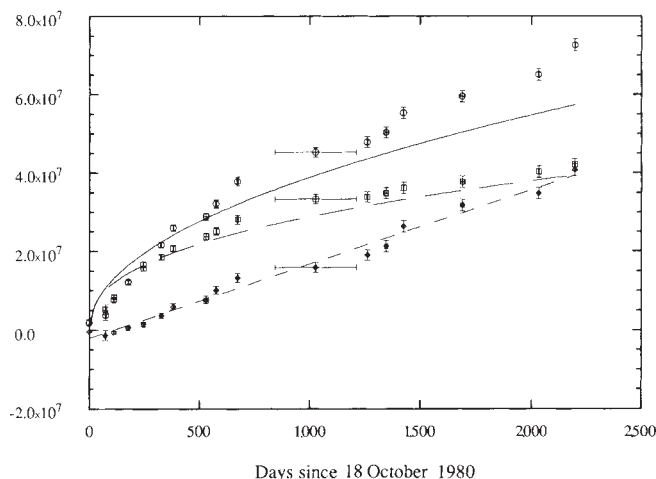


FIG. 1 Total (○), exogenous (□) and endogenous (◇) cumulative dome volumes as a function of the number of days since 18 October 1980. The growth curves are: total volume (solid line) $1.31 \times 10^6 t^{0.49}$, $r^2 = 0.984$; exogenous volume (long dashed line) $1.88 \times 10^6 t^{0.40}$, $r^2 = 0.980$; endogenous volume (short dashed line) $-2.01 \times 10^6 + 1.89 \times 10^5 t$, $r^2 = 0.986$. Negative values for the first two endogenous volumes are caused by mis-registration of the second map, and lack of knowledge of the pre-eruption topography of the crater floor. Horizontal error bars correspond to the duration of individual eruptive episodes. Vertical error bars show uncertainties of 7–9% related to inaccuracies in photogrammetry, digitization and transcription of lobe outlines from air photographs to topographic maps.

According to this view, the total volume production rate is dictated by magma-chamber processes, such as the crystallization of anhydrous minerals. The long-term decrease in overall production suggests that these deep-seated processes became progressively less efficient from 1980 to 1986. Although the partitioning of lava into exogenous and endogenous components is immediately controlled by the thickness and strength of the dome's brittle crust, it is ultimately the lengthening of repose intervals (or some other deep-seated factor) that determines when endogenous inflation can give way to exogenous extrusion.

The actual path taken by the magma to the surface after the crust starts to fail will be controlled by the dome's internal temperature distribution, which in turn reflects details of the emplacement history. After an extrusive episode ends, the fractured pathway to the surface anneals and the entire crust continues to grow inward. Analysis of magnetic data suggests that, overall, this crust thickens linearly and cumulatively with time throughout the dome's history¹⁷. As its strength thus increases, each subsequent intrusive event must exert more force to fracture the crust sufficiently to permit extrusion. As the magma production rate beneath the volcano decreases (as seen in Fig. 1), a progressively larger fraction of each eruptible volume is required to expand the dome before the crust fails, and less is available to form the exogenous lobe. Ultimately, of course, this relationship breaks down when the dome is entirely solid.

This model has several implications for dome emplacement processes. The surprising regularity of endogenous growth, independent of detailed variations in the history and geometry of the dome, indicates that it must be controlled by a regular and continuous process that changes little over the course of a dome's evolution. If healing and thickening of the crust are the causes of this regularity, then they must be insensitive to, or occur at some depth beyond, the influence of near-surface inhomogeneities in the dome, and/or transient processes (such as explosive eruptions). This may imply a single strong zone in the dome interior that serves as a cork in the eruptive system, controlling whether or not magma can break its way to the dome surface. Once this strong area breaks, the magma will follow the easiest path upward. Such an internal structure is analogous to that of the lithosphere, the strength of which increases downward to a maximum value owing to closing of fractures, but then decreases rapidly at greater depths owing to thermal softening.

If solidification and strengthening of the dome interior occur more rapidly than the generation of buoyancy in the magma chamber, it may finally become so difficult for new magma to reach the surface that eruption temporarily shuts down. This might appear to be the end of an eruptive cycle. At such times, the inability of the deep-seated system to release the continuously generated magma overpressure may set the stage for an explosive eruption that clears out the vent at a later date. Other data, for example, SO₂ production¹⁸ or isotope characteristics of magmatic volatiles^{12,13}, might offer clues that degassing was complete.

Our model provides several new conclusions about dome emplacement at Mount St Helens that might be applicable to other volcanoes. First, the predictability of endogenous growth shows that the cooling dome controls its own style of development, independent of such deep-seated factors as magma overpressure and extrusion rate. Second, the regular decrease in exogenous growth rate also allows volume prediction, in this case of the size of a future extrusive lobe based on the repose interval that precedes its appearance. Knowing this volume, one can then determine when an ongoing eruptive event should end. Finally, the observed transition from predominantly exogenous to predominantly endogenous growth reflects the increase in crust thickness, which in turn seems to depend on long repose periods rather than some fundamental change in the character of the dome. These conclusions add to previously reported relationships for total dome growth⁹.

We cannot say at present whether the long-term supply rate and partitioning trends identified in the growth of the Mount St Helens dome apply to other extrusions in different climatic or tectonic settings, or with different magma compositions or supply rates, because no comparable data sets exist. As part of the monitoring programs for active domes like those at Redoubt Volcano in Alaska, Bezymianny in Kamchatka and Santiaguito in Guatemala, vertical aerial photographs suitable for generating topographic maps should be collected on a regular basis. □

Received May; accepted September 1990.

1. Huppert, H. E., Shepherd, J. B., Sigurdsson, H. & Sparks, R. S. J. *J. Volcan. geotherm. Res.* **14**, 199–222 (1982).
2. Taylor, G. A. M. *Aust. Bur. Miner. Res. Bull.* **38**, 1–129 (1958).
3. Swanson, D. A. et al. *Spec. Pap. geol. Soc. Am.* **212**, 1–16 (1987).
4. Rose, W. I., Pearson, T. & Bonis, S. *Bull. Volcan.* **40**, 53–70 (1977).
5. Fink, J. H. & Manley, C. R. *Int. Ass. Volcan. Chem. Earth's Inter. Proc. Volcan. Vol. 1* (ed. Latter, J.) 169–179 (Springer, Berlin, 1989).
6. Heiken, G. & Wohletz, K. *Spec. Pap. geol. Soc. Am.* **212**, 55–76 (1987).
7. Mellors, R. A., Waitt, R. B. & Swanson, D. A. *Bull. Volcan.* **50**, 14–25 (1988).
8. Holcomb, R. T. & Colony, W. E. (abstr.) *Proc. Gen. Assembly Int. Un. Geodesy Geophys. vol. 2*, 417 (IUGG, Vancouver, 1987).
9. Swanson, D. A. & Holcomb, R. T. *Int. Ass. Volcan. Chem. Earth's Inter. Proc. Volcan. Vol. 2* (ed. Fink, J. H.) 1–24 (Springer, Berlin, 1990).
10. Hulme, G. *Geophys. J. R. astr. Soc.* **39**, 361–383 (1974).
11. Cashman, K. V. *Bull. Volcan.* **50**, 194–209 (1988).
12. Anderson, S. W. & Fink, J. H. *Int. Ass. Volcan. Chem. Earth's Inter. Proc. Volcan. Vol. 2* (ed. Fink, J. H.) 25–46 (Springer, Berlin, 1990).
13. Anderson, S. W. & Fink, J. H. *Nature* **341**, 521–523 (1989).
14. Chadwick, W. W. Jr, Archuleta, R. J. & Swanson, D. A. *J. geophys. Res.* **93**, 4351–4366 (1988).
15. Denlinger, R. *Int. Ass. Volcan. Chem. Earth's Inter. Proc. Volcan. Vol. 2* (ed. Fink, J. H.) 70–87 (Springer, Berlin, 1990).
16. Iverson, R. *Int. Ass. Volcan. Chem. Earth's Inter. Proc. Volcan. Vol. 2* (ed. Fink, J. H.) 47–69 (Springer, Berlin, 1990).
17. Dzurisin, D., Denlinger, R. P. & Rosenbaum, J. G. *J. geophys. Res.* **95**, 2763–2780 (1990).
18. Casadevall, T. et al. *Science* **221**, 1383–1385 (1983).

ACKNOWLEDGEMENTS. We thank R. Holcomb, W. Colony and D. Swanson for access to their high-resolution topographic maps. We also thank D. Swanson and R. Denlinger for reviews. This work was supported by grants from the NSF and NASA.

Oxygen fugacity controls in the Earth's upper mantle

C. Ballhaus, R. F. Berry & D. H. Green

Department of Geology, University of Tasmania, Hobart, Tasmania 7001, Australia

DESPITE significant progress in recent years^{1–3}, there is still considerable debate surrounding the oxygen fugacity (f_{O_2}) of the Earth's upper mantle and the chemical reactions that buffer it. Electrochemical measurements^{1,3} give heterogeneous oxidation states, with one group near the fayalite–magnetite–quartz (FMQ), and the other near the iron–wüstite buffer (IW). Thermobarometric calculations based on Fe³⁺/Fe²⁺ ratios in basaltic glasses⁴, mantle-derived spinels^{5,6}, ilmenites⁷ and garnets⁸ support uniform oxidation states bounded by wüstite–magnetite (WM) and FMQ. Here we compare the behaviour of four 'oxygen barometers'^{5,6,9,10} based on the compositions of coexisting minerals, including a new empirical olivine–orthopyroxene–spinel barometer¹⁰ that has been calibrated experimentally in spinel lherzolite at pressures, temperatures and compositions appropriate to the upper mantle. Application of this barometer to mantle-derived rocks and basaltic melts of different tectonic settings suggests that the upper mantle is only weakly buffered by Fe³⁺/Fe²⁺ equilibria. Large-scale heterogeneity in f_{O_2} structure is governed by recycling processes and injection of oxidized crustal material into a moderately reduced, poorly buffered asthenosphere.

The ferric iron content in spinel coexisting with olivine and orthopyroxene forms the basis for several recent oxygen barometers^{5,6,9,10} applicable to mantle-derived rocks. O'Neill and Wall⁵ derived a solution model for the activity of magnetite in Cr–Al spinel from published thermodynamic data. Mattioli and Wood⁶ calibrated the activity of magnetite by measuring

intrinsic f_{O_2} values of synthetic $MgAl_2O_4$ - Fe_3O_4 solid solutions in an electrochemical cell. Wood *et al.*⁹ presented a model based on high-temperature thermopower measurements of order-disorder in spinel. Recognizing the need for experimental data at high Cr/(Cr + Al) ratios, especially at upper-mantle pressures, we equilibrated synthetic Cr-Al spinel lherzolite and harzburgite assemblages over a range of pressure ($P = 1$ –2.5 GPa), temperature ($T = 1,313$ –1,573 K), f_{O_2} (iron-wüstite to magnetite-haematite) and mole fractions of Cr ($X = 0.19$ –0.85) compatible with upper-mantle conditions¹⁰. We have formulated an empirical oxygen sensor that is applicable to Cr-Al spinels in mantle-derived rocks and primitive basaltic melts

$$\begin{aligned} \Delta \log (f_{O_2}) = & 0.27 + 2,505/T - 400P/T - 6 \log (X_{Fe}^{olv}) \\ & - 3,200(1 - X_{Fe}^{olv})^2/T + 2 \log (X_{Fe}^{sp,2+}) \\ & + 4 \log (X_{Fe}^{sp,3+}) + 2,630(X_{Al}^{sp})^2/T \end{aligned}$$

where f_{O_2} is calculated relative to the FMQ buffer¹¹, T is in K, P is in GPa, (X_{Fe}^{olv}) and ($X_{Fe}^{sp,2+}$) are the $Fe^{2+}/(Mg + Fe^{2+})$ cation ratios in olivine and spinel, and ($X_{Fe}^{sp,3+}$) and (X_{Al}^{sp}) are the $Fe^{3+}/\Sigma R^{3+}$ and $Al/\Sigma R^{3+}$ cation ratios in spinel (ΣR^{3+} is the total content of trivalent ions).

In Fig. 1 we apply these oxygen barometers to 170 analyses from a worldwide sampling of spinel peridotite xenoliths. The O'Neill-Wall⁵ model (Fig. 1a) gives an f_{O_2} range of about 2.5 log units and seems to correct well for non-ideal Cr-Al mixing, assuming that there is no systematic change in oxidation state across the range of Cr/(Cr + Al) in this suite. Mattioli and Wood's⁶ version (Fig. 1b) is more sensitive to $Fe^{3+}/\Sigma Fe$ variations and gives at least twice the range in f_{O_2} , with many spinels approaching and plotting below IW. The recent version of Wood *et al.*⁹ (Fig. 1c) and our empirical model¹⁰ (Fig. 1d) imply systematic oxidation with increasing Cr content in these samples.

These differences in redox trends show that inferences of redox conditions in the upper mantle are very model-dependent. For example, is mantle metasomatism coupled with oxidation

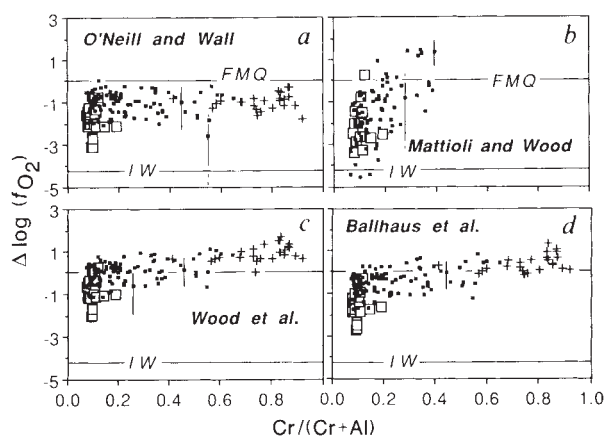


FIG. 1 Application of the four oxygen barometers^{5,6,9,10} to 170 spinel peridotite xenoliths worldwide (f_{O_2} relative to FMQ). □, 'primitive' (least metasomatized³²) lherzolite xenoliths with flat light-rare-earth-element patterns; ■, lightly metasomatized lherzolites and harzburgites with light-rare-earth-element enrichment and occasional pargasite; +, strongly metasomatized samples with metamorphic Ti-phlogopite and kaersutite and some Ti and Mn enrichment in spinel. All calculations for 1.5 GPa and olivine-spinel temperatures¹⁰ (~900–1,100 °C). FMQ buffer ($3Fe_2SiO_4 + O_2 \rightleftharpoons 2Fe_3O_4 + 3SiO_2$) and IW buffer ($2Fe + O_2 \rightleftharpoons 2FeO$) from ref. 11. IW buffer position calculated for 1,000 °C. Error bars assume a generous analytical error in microprobe analyses of 1.5 wt% relative, with cumulative effect on calculated ferric iron (for example -1.5% RO, +1.5% R_2O_3 in spinel). Mattioli and Wood⁶ version only applied where $X_{Cr}^{spinel} \leq 0.4$, because of the limited range of its calibration.

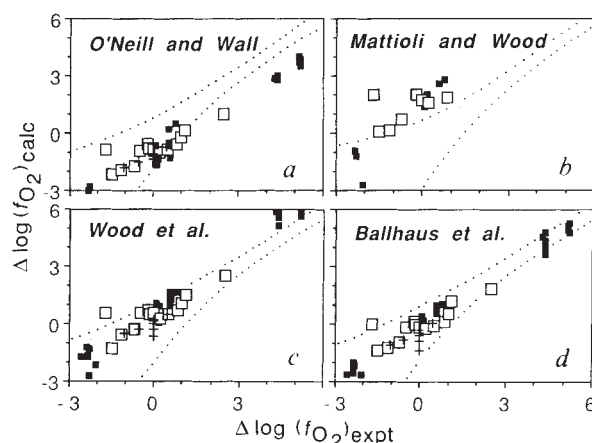


FIG. 2 Calculated f_{O_2} plotted against experimental f_{O_2} (relative to FMQ) for synthetic spinel compositions^{10,14,15}. Region embraced by dotted lines illustrates errors in calculated f_{O_2} , calculated as in Fig. 1. The experimental data of ref. 10 are not independent and are only included to highlight systematic differences among the calibrations at high f_{O_2} . Spinel with $Cr/(Cr + Al) > 0.4$ not plotted with the Mattioli-Wood⁶ calibration.

as implied by our empirical model (Fig. 1d) and the barometer of Wood *et al.* (Fig. 1c), or not (as implied by the O'Neill-Wall⁵ model in Fig. 1a)? Is the continental lithosphere as heterogeneous with respect to f_{O_2} as the Mattioli-Wood⁶ model implies (Fig. 1b)? If the Cr-Al correction of the O'Neill-Wall⁵ solution model is accurate then upper-mantle metasomatism by carbonatite melts¹² could be excluded, as all xenolith samples plot below carbonate stability¹³ (about FMQ -0.2 log units at 1,000 °C and 1.5 GPa). The oxidation trend implied in Fig. 1c, d, on the other hand, is consistent with carbonate stability and influx of carbonatite melts.

To assess which model is the most accurate, we apply each oxygen barometer to experimental spinel compositions synthesized under conditions of controlled f_{O_2} ^{10,14,15} (Fig. 2), and to spinel phenocrysts in basaltic melts, the oxidation states of which are constrained by the measured $Fe^{3+}/\Sigma Fe$ bulk-rock ratios⁴ (Fig. 3). The O'Neill-Wall model⁵ underestimates experimental oxygen fugacities (Fig. 2a) by ~0.5–1 log units below FMQ, and ~1.5 log units at higher f_{O_2} . Relative to whole-rock oxidation states, the model tends to compress the f_{O_2} values calculated from spinels, such that a discrimination between the groups of basalts on the basis of f_{O_2} becomes impossible (Fig. 3a). An assessment of the Mattioli-Wood oxygen barometer⁶ is impeded by its narrow range of application (see Fig. 1). The experimental and natural data that fall inside its limits (Figs. 1b, 2b) indicate that the Mattioli-Wood model is too sensitive to small variations in Fe^{3+} and poorly corrected for Cr-Al interaction; it creates large apparent f_{O_2} heterogeneities in data that are essentially the same within analytical error (the 'mantle average array' in ref. 16). The best overall fit is achieved with our empirical model¹⁰ and the barometer of Wood *et al.*⁹, although the latter model overestimates experimental f_{O_2} (Fig. 2c) and whole-rock Fe^{3+}/Fe^{2+} f_{O_2} (Fig. 3b) by 0.5–1 log units.

The close agreement between two independent studies^{9,10} suggests that oxygen thermobarometry based on spinel compositions is sufficiently well advanced to give results that are relevant to the redox state of the upper mantle. In Fig. 4 we apply our empirical model to mantle-derived rocks and basaltic melts, in addition to the xenolith suite shown in Fig. 1d. The following generalizations are apparent. (1) Modern sub-continental and sub-oceanic fertile lithosphere and asthenosphere ('primitive' peridotite xenoliths, abyssal spinel peridotites, mid-ocean-ridge basalt (MORB) source regions) are moderately reduced from

around FMQ-2 to FMQ, with no significant differences between sub-continental and sub-oceanic upper mantle. Low-pressure magnesiochromite phenocrysts in MORB are more oxidized than coexisting high-pressure Al-spinel megacrysts^{14,17}, but both spinel populations fall within the f_{O_2} range calculated from Fe^{3+}/Fe^{2+} ratios of MORB whole-rock and glass compositions¹⁸. A large-scale f_{O_2} heterogeneity proposed for sub-oceanic lithosphere¹⁹ is not supported. (2) Geochemical enrichment processes in the upper mantle are oxidizing. Enriched MORB²⁰, metasomatized spinel lherzolite xenoliths from continental rifts, 'mantle-plume'-related ocean island basalts²¹ (OIBs) and lithosphere affected by 'plumes' are oxidized to around FMQ to FMQ+2. The enriched component is at least as oxidized as FMQ+1, accepting the most metasomatized xenoliths in Fig. 1d as a lower limit. (3) Upper mantle modified by subduction, as represented by island arc basalts (IABs) and their cumulate nodules, is highly oxidized with f_{O_2} between FMQ+1 and FMQ+3.

It can be debated whether spinel phenocrysts in basalts are reliable indicators of the redox states of the basalt source regions, or whether their redox state has been shifted relative to their mantle sources by degassing^{22,23} or by changes in Fe^{3+}/Fe^{2+} liquid and spinel ratios on decompression²⁴. We believe that syn- or post-eruptive f_{O_2} changes are not substantial. Where we do have the additional redox information from associated upper-mantle xenoliths (spinel lherzolite inclusions in OIBs), from coexisting high-pressure (upper mantle) aluminous spinel megacrysts in MORB, or from high-pressure (lower crust/upper mantle) cumulate nodules in IABs, the low-pressure spinels are only slightly more oxidized than the mantle samples. Thus, there is evidence for differences between the f_{O_2} of basalts and their sources, of the order of +0.5 to 1 log units, but the discrepancies are not large enough to alter our conclusions.

Our results point to a strong correlation between geotectonic setting and redox state, and impose constraints on redox control in the upper mantle. An important unresolved question is whether f_{O_2} control is provided by (1) single buffer reactions such as carbonate-graphite-silicate (EMOG)¹³ or graphite-CO-CO₂±H₂O fluid equilibria (CCO buffer)²²; (2) 'open-system buffering' through pervasive C-H-O degassing^{2,25}; or (3) sliding Fe^{3+}/Fe^{2+} equilibria¹. An f_{O_2} range of roughly four log units in modern upper mantle cannot be reconciled with a single buffer reaction. Primitive and some lightly metasomatized xenoliths do fall in the general range of EMOG¹² but enriched xenoliths,

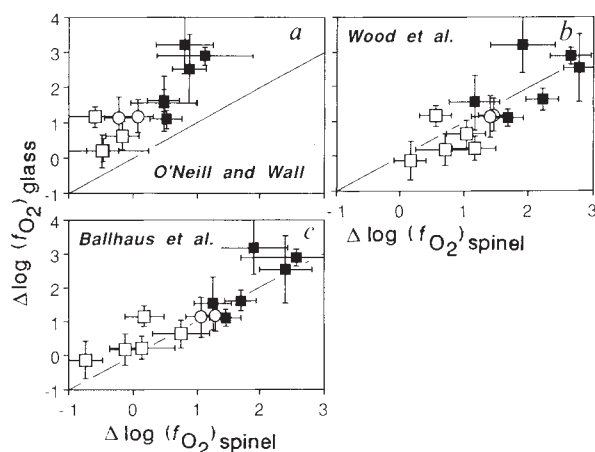


FIG. 3 Plot of f_{O_2} calculated from bulk-rock Fe^{3+}/Fe^{2+} ratios⁴ against f_{O_2} calculated from spinel compositions (relative to FMQ; error bars are two standard deviations on analytical ranges). □, mid-ocean-ridge basalts; ○, ocean island basalts; ▴, island arc basalts and andesites. All spinels are more Cr-rich, $Cr/(Cr+Al) > 0.4$, and outside the range of the Mattioli-Wood⁶ oxygen barometer.

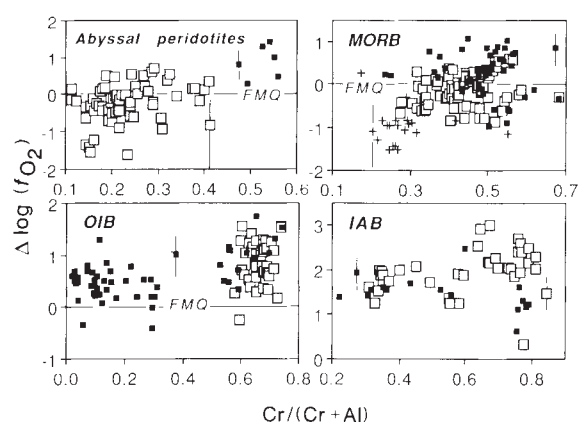


FIG. 4 Oxygen fugacities calculated¹⁰ for spinels in abyssal spinel (□) and plagioclase peridotites (■) magnesiochromite phenocrysts in depleted (□) and enriched (■) mid-ocean-ridge basalts (N-MORB and E-MORB) and high-pressure^{14,17} Al-spinel megacrysts in MORB (+), spinels in ocean island basalts (OIBs, □) and their mantle xenoliths (■), and spinels in island arc basalts (IABs, □) and their cumulate nodules (■). Assumed pressures are 1.5 GPa for mantle-derived rocks and 0.1 GPa for liquidus spinels in basaltic melts. Absence of orthopyroxene in MORBs, OIBs and some IABs does not significantly affect calculated f_{O_2} (see ref. 10). Representative error bars calculated as in Fig. 1. A list of references to the data in Figs 1, 3 and 4 is available from the authors.

OIBs and especially IABs are incompatible with graphite and well within the carbonate-CO₂ stability field. Our data support the third alternative, that the oxygen fugacity of the upper mantle is weakly buffered by sliding ferric/ferrous redox equilibria, the buffering capacity of which depends on the pressure, temperature and bulk Fe^{3+}/Fe^{2+} ratio. A simple mass balance calculation using our empirical oxygen sensor¹⁰ shows that an influx of 25 p.p.m. oxygen under isothermal and isobaric conditions into depleted reduced residual uppermost-mantle spinel harzburgite¹⁷ (bulk $Mg/(Mg+Fe) = 0.9$) at FMQ-2 would shift the f_{O_2} by two log units, whereas the same amount of oxygen added to more oxidized harzburgite at FMQ+1 would shift the f_{O_2} by less than 0.3 log units (assuming 90% of ferric iron dissolves as magnetite in spinel). Thus, Fe^{3+}/Fe^{2+} equilibria do provide effective f_{O_2} buffering above ~FMQ+1, but are inefficient below FMQ. The apparent sensitivity of primitive lithosphere to adjust to oxidizing metasomatizing fluids (see the oxidation trend in Fig. 1d) suggests that reduced upper mantle is poorly buffered; graphite in particular cannot be generally present. On the other hand, if subduction has indeed been active since Archaean times, yet has not oxidized MORB source regions to $f_{O_2} > FMQ$, there may well be a role for continued degassing of reduced (CH₄-H₂O-H₂) fluids from the core or deep mantle.

The observation that OIBs and metasomatized sub-continental and sub-oceanic lithosphere samples are oxidized, is evidence for recycling of a crustal 'hydrophile' component into the upper mantle. Given the highly oxidized nature of primitive IABs and their cumulate nodules, the likely source for 'excess oxygen' in these samples is subducted altered oceanic crust, sediments or recycled mantle wedge that has been oxidized above the dehydrating slab. Following entrainment into the upper mantle, oxidized geochemically enriched well buffered entities of subducted material may accumulate in the transition zone^{2,26}. Redox reactions² and/or conductive heating of subducted material to ambient asthenosphere temperature will trigger partial melting^{2,27}, and small highly enriched oxidized melt fractions will eventually be released into overlying upper mantle to feed mantle plumes or to cause mantle metasomatism. The preservation of variably oxidized domains in the deep asthenosphere is made possible by a combination of factors: (1) low self-buffering capacity of reduced ambient asthenosphere with

respect to f_{O_2} ; (2) limited mixing between oxidized recycled material and reduced asthenosphere, and long-lived preservation of separate mantle reservoirs (supported by isotope evidence²¹); (3) inability of a separate C-H-O fluid phase in the ambient asthenosphere to cross the discontinuity in f_{O_2} without triggering melting and dissolving in a melt. Our data do not support the generalization that lithosphere is more reduced than asthenosphere²⁸ (based on the occurrence of diamonds in thick, old lithosphere²⁹ and measurements of intrinsic oxygen fugacity³) in the modern mantle. Instead, we interpret highly reduced diamond-bearing lithosphere³⁰ as a relict from an earlier lower- f_{O_2} regime in the outer layers of the primitive Earth, possibly determined by CH_4 - H_2O - H_2 deep-mantle degassing^{1,2,31}. □

Received 24 August; accepted 9 October 1990.

1. Arculus, R. J. *A. Rev. Earth planet. Sci.* **13**, 75-95 (1985).
2. Green, D. H., Falloon, T. J. & Taylor, W. R. in *Magmatic Processes* (ed. Mysen, B. O.) 139-154 (Geochem. Soc., Lond., 1987).
3. Arculus, R. J., Dawson, J. B., Mitchell, R. H., Gust, D. A. & Holmes, R. D. *Contr. Miner. Petrol.* **85**, 85-94 (1984).
4. Kilinc, A., Carmichael, I. S. E., Rivers, M. L. & Sack, R. O. *Contr. Miner. Petrol.* **83**, 136-140 (1983).
5. O'Neill, H. St. C. & Wall, V. J. *J. Petrol.* **28**, 1069-1102 (1987).

6. Mattioli, G. S. & Wood, B. J. *Contr. Miner. Petrol.* **98**, 148-162 (1988).
7. Eggler, D. H. *Geophys. Res. Lett.* **10**, 365-368 (1983).
8. Luth, R. W., Virgo, D., Boyd, F. R. & Wood, B. J. *Contr. Miner. Petrol.* **104**, 56-72 (1990).
9. Wood, B. J., Bryndzia, L. T. & Johnson, K. E. *Science* **248**, 337-345 (1990).
10. Ballhaus, C., Berry, R. F. & Green, D. H. *Contr. Miner. Petrol.* (in the press).
11. O'Neill, H. St. C. *Am. Miner.* **72**, 67-75 (1987).
12. Green, D. H. & Wallace, M. E. *Nature* **336**, 459-462 (1988).
13. Eggler, D. H. & Baker, D. R. in *High Pressure Research in Geophysics* (eds Akimoto, S. & Manghnani, M. H.), 237-250 (1982).
14. Fisk, M. R. & Bence, A. E. *Earth planet. Sci. Lett.* **48**, 111-123 (1980).
15. Murck, B. W. & Campbell, I. H. *Geochim. cosmochim. Acta* **50**, 1871-1887 (1986).
16. Mattioli, G. S., Baker, M. G., Rutter, M. J. & Stolper, E. M. *J. Geol.* **97**, 521-536 (1989).
17. Green, D. H., Hiberson, W. O. & Jaques, A. L. in *The Earth* (ed. McElhinney, M. W.), 265-299 (Academic, London, 1979).
18. Christie, D. M., Carmichael, I. S. E. & Langmuir, C. H. *Earth planet. Sci. Lett.* **79**, 397-411 (1986).
19. Bryndzia, L. T., Wood, B. J. & Dick, H. J. B. *Nature* **341**, 526-527 (1989).
20. Sun, S. S., Nesbitt, R. W. & Sharaskin, A. Y. *Earth planet. Sci. Lett.* **44**, 119-138 (1979).
21. Sun, S. S. & McDonough, W. F. in *Magnetism in the Ocean Basins* (eds Saunders, A. D. & Norry, M. J.) 313-345 (Blackwell, Oxford, 1989).
22. Sato, M. *Geophys. Res. Lett.* **5**, 447-449 (1978).
23. Mathez, E. A. *Nature* **310**, 371-375 (1984).
24. Gudmundsson, G. & Holloway, J. R. *Terra Abstr.* **2**, 14 (1990).
25. Woermann, E. & Rosenhauer, M. *Fortschr. Miner.* **83**, 263-349 (1985).
26. Ringwood, A. E. *J. Geol.* **90**, 611-643 (1982).
27. Ringwood, A. E. *Chem. Geol.* **82**, 187-207 (1990).
28. Haggerty, S. E. & Tompkins, L. A. *Nature* **303**, 295-300 (1983).
29. Richardson, S. H., Gurney, J. J., Erlank, A. J. & Harris, J. W. *Nature* **310**, 198-202 (1984).
30. Haggerty, S. E. *Nature* **320**, 34-39 (1986).
31. Taylor, W. R. & Green, D. H. *Nature* **332**, 349-352 (1988).
32. Jochum, K. P., McDonough, W. F., Palme, H. & Spettel, B. *Nature* **340**, 548-550 (1989).

ACKNOWLEDGEMENTS. We thank K. L. Harris, S. Stephens and W. Jablonski for technical support. Supported by the German and Australian Research Councils.

Alteration in crossbridge kinetics caused by mutations in actin

Douglas R. Drummond, Michelle Peckham,
John C. Sparrow & David C. S. White

Department of Biology, University of York, York YO1 5DD, UK

THE generation of force during muscle contraction results from the interaction of myosin and actin. The kinetics of this force generation vary between different muscle types and within the same muscle type in different species^{1,2}. Most attention has focused on the role of myosin isoforms in determining these differences³⁻⁵. The role of actin isoforms has received little attention, largely because of the lack of a suitable cell type in which the myosin isoform remains constant yet the actin isoforms vary. An alternative approach would be to examine the effect of actin mutations, however, most of these cause such gross disruption of muscle structure that mechanical measurements are impossible⁶⁻¹⁰. We have now identified two actin mutations which, despite involving conserved amino acids, can assemble into virtually normal myofibrils¹¹. These amino-acid changes in actin significantly affect the kinetics of force generation by muscle fibres. One of the mutations is not in the putative myosin-binding site, demonstrating the importance of long-range effects of amino acids on actin function.

We used *in vitro* mutagenesis to produce single amino-acid changes in the product of *Act88F*, the only actin gene expressed in the indirect flight muscles of *Drosophila melanogaster*¹², and reintroduced the mutant genes into flies homozygous for the *Act88F* null mutation *KM88* using P-element transformation¹¹. Two mutants *E316K* (lysine replaces glutamic acid at amino acid 316) and *G368E* (glutamic acid replaces glycine at amino acid 368) gave muscles with apparently normal structure when examined by light microscopy, and with only slight disruption of their ultrastructure. To confirm that the transformed flies were expressing the mutant actin, muscles were dissected from the mutant and wild-type flies and analysed by two-dimensional gel electrophoresis (Fig. 1). This showed that only the mutant actin was present. Densitometry of mutant and wild-type muscle proteins separated by one-dimensional SDS-PAGE showed similar myosin/actin ratios, confirming that the transformed lines of both *E316K* and *G368E* flies have normal levels of

TABLE 1 Mechanical properties of indirect flight muscles

	Wild type	G368E	E316K	KM88/+
Rate constants of the phases in the active response				
r_2	4,488±448 (14)	3,260±319 (14)	*7,149±758 (11)	*2,242±483 (6)
r_3	241±14 (15)	*161±14 (16)	*329±26 (11)	202±16 (8)
r_4	22±3 (15)	15±1 (16)	21±2 (11)	15±2 (8)
Tension (pN per A-filament)				
Active	19±3 (18)	*29±3 (15)	24±2 (11)	*33±3 (8)
Rigor	32±4 (15)	39±7 (12)	25±4 (8)	45±5 (7)
Stiffness (pN per A-filament per % strain)				
Relaxed	35±4 (17)	43±5 (12)	41±7 (8)	35±3 (8)
Active	53±5 (18)	*84±7 (12)	65±8 (11)	60±5 (8)
Rigor	215±26 (15)	242±43 (14)	159±14 (11)	188±26 (8)

The time course of the tension response (Fig. 2) was analysed as a sum of exponentials ($A_1 + A_2 e^{-r_2 t} + A_3 e^{-r_3 t} + A_4 e^{-r_4 t}$), and the rate constants for the three phases 2, 3 and 4 (r_2 , r_3 and r_4) shown in Fig. 2 are given. The steady state tension and stiffness measurements, normalization by A-filament content, and the solution compositions were as described previously²². Values are given as the mean ± s.e.m. The numbers of fibres tested are shown in brackets.

* Significant difference ($P < 0.05$) from wild type.

actin. Flight testing of the flies showed that *E316K* flies were as flightless as the *Act88F* null strain *KM88*, whereas *G368E* flies could fly, although less well than wild-type flies¹¹.

To test the effect of these mutations on the other properties of the flight muscle, we measured the wing-beat frequency of the flies in flight (if they flew), the lattice packing of the muscle, and the mechanics of the demembranated muscle fibres in relaxing, activating and rigor solutions.

The wing-beat frequency of 2-3-day-old female wild-type and *G368E* flies was measured at 18 °C and 59% relative humidity in free flight, using an optical tachometer¹³. The wingbeat frequency of the *G368E* flies (178 ± 3 Hz, mean ± s.e.m., 23 flies tested) was reduced significantly, by about 20%, compared to wild type (227 ± 4 Hz, 29 flies tested). *E316K* flies did not produce a continuous wing beat and could not be tested.

Both the *E316K* and *G368E* mutations altered the mechanics of demembranated muscle fibres compared to those of wild type (Table 1). In *G368E*, the rate constant r_3 for the increase in delayed tension which followed a quick stretch in activating conditions (phase 3, Fig. 2) was reduced significantly by about 30% from 241 s^{-1} to 161 s^{-1} ($P < 0.05$, Student's *t*-test, Fig. 2; Table 1). The rate constants for tension relaxation during phases

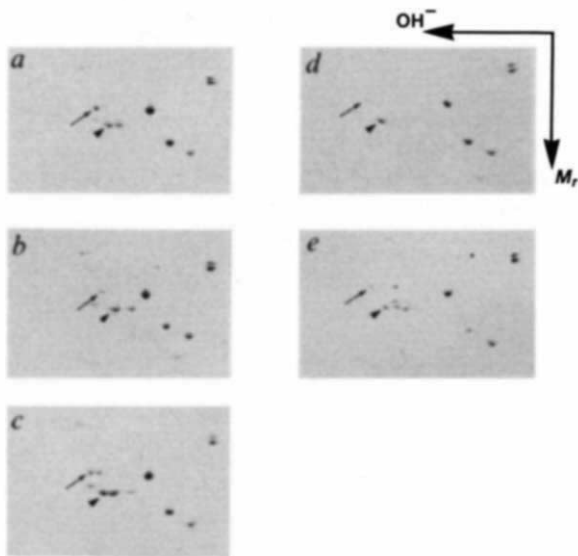


FIG. 1 The indirect flight muscles were dissected from *a*, wild-type, *b*, *G368E*, *c*, a mixture of wild-type and *G368E*, *d*, *E316K* and *e*, a mixture of wild-type and *E316K* flies; the muscle proteins were separated by two-dimensional PAGE²³ and the gels stained with Coomassie blue. The *G368E* (*b*) and *E316K* (*d*) flies have two actin (large arrowhead) and one arthrin (small arrowhead) spots as found in wild-type flies (*a*). Only the *Act88F* gene is expressed in indirect flight muscles and the actin protein is post-translationally modified by the *mod* gene product to give two actin spots²⁴, and by conjugation of ubiquitin to form arthrin¹². In *G368E* flies all the actin moves to a more acidic position relative to the other proteins and in *E316K* flies to a more basic position, as would be expected from the charge change caused by each mutation. This was confirmed by mixing indirect flight muscle proteins from the mutant and wild-type flies (*c*, *e*). The mutants *E316K* and *G368E* were created *in vitro* by mutation of the *Act88F* gene as described¹¹. The complete mutant *Act88F* gene was then inserted into the genome of a strain which does not express any actin in the indirect flight muscles (*ry*⁵⁰⁶ *KM88 e*^s strain of *Act88F* null flies²⁵) by P-element transformation²⁶. The flies were made homozygous by standard genetical methods and lines which had normal myosin/actin ratios, indicating correct expression levels of the inserted actin gene, were used. *M_r*, relative molecular mass.

2 and 4 (r_2 and r_4) were also reduced (Fig. 2, Table 1) although the differences were not significant. This may be because the determination of the rate constants for phases 2 and 4 is subject to greater experimental error. The amplitudes of phases 2, 3 and 4 were the same as for wild type (data not shown). The steady-state active tension and stiffness were significantly increased (Table 1). The rigor tension and stiffness were increased but not significantly (Table 1). However, the data for the mutant was more variable than that for the wild type. The relaxed stiffness was unchanged.

In *E316K* the rate constant for the increase in delayed tension which followed a quick stretch in activating conditions (phase 3, Fig. 2) was increased by about 30% from 241 s⁻¹ to 329 s⁻¹ ($P < 0.05$, Student's *t*-test, Fig. 2, Table 1). The rate constant for tension relaxation during phase 2 was also significantly increased (Fig. 2, Table 1). The amplitudes of phases 2, 3 and

4 were the same as wild type (data not shown). Steady-state tension and stiffness in active, rigor and relaxed fibres were unchanged.

As there is some disruption of the ultrastructure of these mutants¹² we examined the lattice packing of the flight muscle by electron microscopy (Fig. 3). In *E316K* occasional small holes appear within the lattice structure, whereas in *G368E* there is disruption of the lattice towards the periphery of the myofibrils. To control for these differences affecting the function of the muscle we have used *KM88* heterozygotes. These flies have only one copy of the wild-type *Act88F* gene. This causes disruption of a much greater proportion of the IFM than in either of the mutants (Fig. 3). The flight ability of these flies is less than that of *G368E* but greater than that of *E316K*.

The mechanical properties of the flight muscles of *KM88* heterozygotes are also different from the mutants (Table 1, Fig.

FIG. 2 The tension responses (lower traces) to a rapid stretch (upper trace) in activating solution for demembranated fibres from the indirect flight muscles of wild-type and mutant (*G368E*, *E316K* and *KM88/+*) flies. The activating solution contained 15 mM MgCl₂, 15 mM ATP, 5 mM Calcium-EGTA, 12 mM KCl, 20 mM histidine; pH 7.0, 15 °C. The tension response was fitted as a sum of three exponentials for phases 2, 3 and 4 as described in Table 1. The fibres were demembranated, and mounted on a mechanical rig as described previously²². In most cases, the procedure was slightly modified in that the fibres were demembranated for a minimum of 15 min in a skinning solution containing 15 mM MgCl₂, 15 mM ATP, 5 mM EGTA, 13 mM KCl, 20 mM histidine, 50 µg ml⁻¹ saponin; pH 7.0, 4 °C, and used immediately.

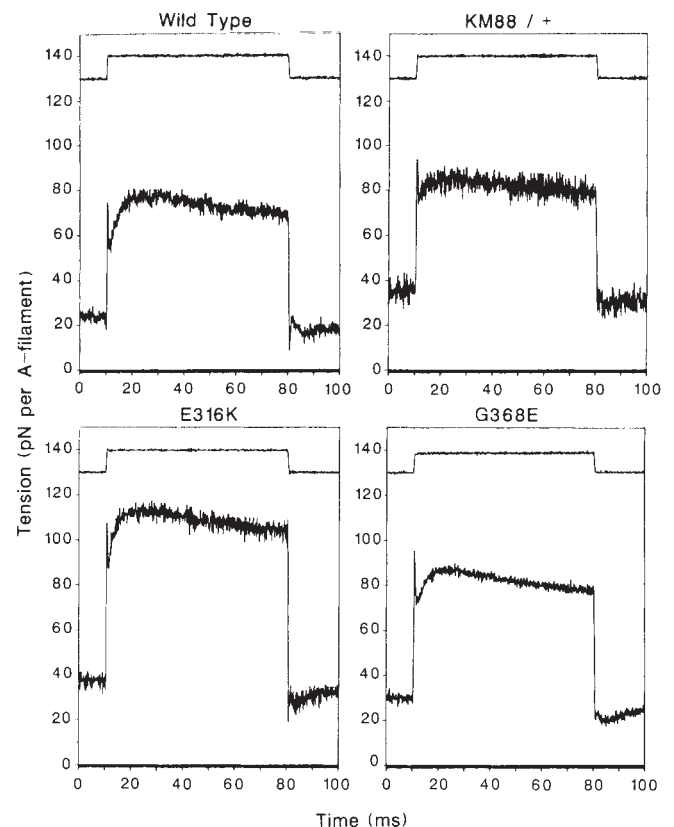
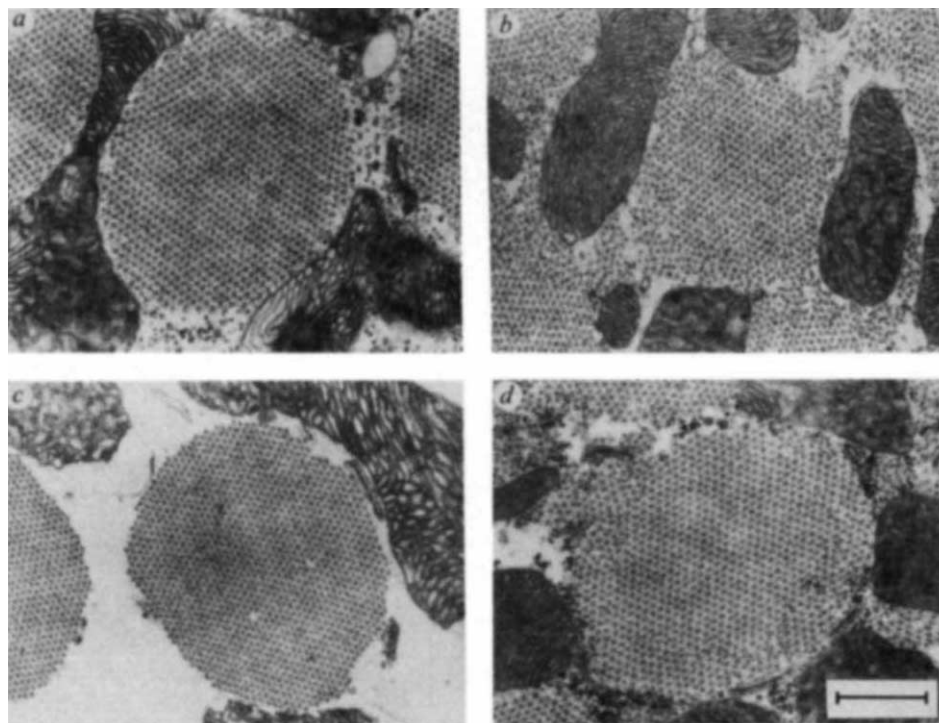


FIG. 3 Electron micrographs showing transverse sections of wild-type (a), *KM88/+* (b), *E316K* (c) and *G368E* (d) indirect flight muscle fibres prepared as in ref. 27. Magnification, $\times 15,000$ scale bar represents 500 nm.



2). In *E316K* the values of r_2 and r_3 , as defined in Table 1, are increased compared to wild type, whereas those of *KM88* heterozygotes are reduced. In *G368E* r_2 and r_3 are also reduced compared to wild type. In contrast to *KM88* heterozygotes the reduction in r_2 is less, whereas the reduction in r_3 is greater in *G368E*. Furthermore, although active tension is increased in both *KM88* heterozygotes and *G368E* compared to wild type, active stiffness is increased significantly only in *G368E*. The qualitative and quantitative differences between the properties of the *KM88* heterozygotes and those of the mutants compared to wild type suggest that these differences cannot simply be explained by different amounts of disruption in the lattice packing of the myofibrils.

The changes in flight ability, wing beat frequency and mechanics of the *E316K* and *G368E* mutants are due solely to the single amino-acid changes in their actins. Furthermore, most of the differences between the mechanical properties of the mutants and wild type cannot be explained by the slight disruption of myofibrillar structure. Curiously, both mutations occur in conserved regions of actin. In all actins amino acid 368 is either glycine or serine, whereas amino acid 316 is always glutamic acid with the exception of a single actin which contains aspartic acid¹⁴. In both mutants the actin appears to be stable and can assemble to form virtually normal myofibrils. Actin is a highly conserved protein^{15,16} and it has been speculated that this results either from the constraints of protein folding¹⁷ or the requirement for actin to interact with many other proteins¹⁸. Our results demonstrate that despite these constraints even nonconservative amino-acid changes can be tolerated in at least some cases.

The results also demonstrate that amino-acid changes in actin can have a significant effect on the mechanical properties of muscle. Amino acid 368 is in the small domain of actin¹⁹ in a region which may be involved in the binding of tropomyosin and the alkali light chain of myosin²⁰. That a change here should affect muscle function is perhaps not surprising. In contrast, amino acid 316 is in the large domain¹⁹, distant from the nearest known myosin or tropomyosin contact at amino acid 336 (ref. 21) suggesting that *E316K* may be affecting the interaction with myosin through a conformational change in the actin. Only by mutating amino acids distant from actual contact sites is it

possible to determine which long range interactions in actin are important for its function.

That single amino-acid changes in actin can affect the kinetics of contraction raises the question of whether the differences which are found in naturally occurring actin isoforms are important in determining the properties of the muscle. Certainly fairly small differences in myosin isoforms have been shown to be functionally significant^{4,5}. *Drosophila* indirect flight muscle provides an ideal system in which to test this directly. Although the mutants examined are clearly defective in terms of powering flight in *Drosophila*, it is not obvious that the same changes in kinetics would be equally disadvantageous in other insects or other muscle types. The particular selection pressures that have maintained the extreme conservation at these amino-acid positions throughout the evolution of actin remain unclear. □

Received 27 July; accepted 5 October 1990.

1. Close, R. I. *Physiol. Rev.* **52**, 129–197 (1972).
2. Molloy, J. E., Kyrtatas, V., Sparrow, J. C. & White, D. C. S. *Nature* **328**, 449–452 (1987).
3. Hoh, J. F. Y., Hughes, S., Hale, P. T. & Fitzsimons, R. B. *J. musc. Res. Cell Motil.* **9**, 30–47 (1988).
4. Schwartz, K. et al. *J. molec. cell. Cardiol.* **13**, 1071–1075 (1981).
5. McNally, E. M., Kraft, R., Bravo-Zehender, M., Taylor, D. A. & Leinwand, L. A. *J. molec. Biol.* **210**, 665–671 (1989).
6. Mogami, K. & Hotta, Y. *Molec. gen. Genet.* **183**, 409–417 (1981).
7. Karlik, C. C., Coutu, M. D. & Fyrberg, E. A. *Cell* **38**, 711–719 (1984).
8. Karlik, C. C., Saville, D. L. & Fyrberg, E. A. *Molec. cell. Biol.* **7**, 3084–3091 (1987).
9. Sakai, Y., Okamoto, H., Mogami, K., Yamada, T. & Hotta, Y. *J. Biochem.* **107**, 499–505 (1990).
10. Reedy, M. C., Beall, C. & Fyrberg, E. *Nature* **339**, 481–483 (1989).
11. Drummond, D. R., Hennessey, E. S. & Sparrow, J. C. *Molec. gen. Genet.* (in the press).
12. Ball, E. et al. *Cell* **51**, 221–228 (1987).
13. Unwin, D. M. & Ellington, C. P. *J. exp. Biol.* **82**, 377–378 (1979).
14. Akrigg, D. et al. *Nature* **335**, 745–746 (1988).
15. Hightower, R. C. & Meagher, R. B. *Genetics* **114**, 315–332 (1986).
16. Pollard, T. D. & Cooper, J. A. *A. Rev. Biochem.* **55**, 987–1035 (1986).
17. Schutt, C. E. & Lindberg, U. in *Molecular Mechanisms in Muscular Contraction* (ed. Squire, J. M.) 49–63 (Macmillan, Basingstoke, UK, 1990).
18. Moir, A. J. G. & Levine, B. A. *J. inorg. Biochem.* **28**, 271–278 (1986).
19. Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. *Nature* **347**, 37–44 (1990).
20. Hamblly, B. D., Barden, J. A., Miki, M. & Dos Remedios, C. G. *Bioessays* **4**, 124–128 (1986).
21. Szilagyi, L. & Lu, R. C. *Biochem. biophys. Acta* **709**, 204–211 (1982).
22. Peckham, M., Molloy, J. E., Sparrow, J. C. & White, D. C. S. *J. Muscle Res. Cell Motil.* **11**, 203–215 (1990).
23. O'Farrell, P. H. *J. biol. Chem.* **250**, 4007–4021 (1975).
24. Mahaffey, J. W., Coutu, M. D., Fyrberg, E. A. & Inwood, W. *Cell* **40**, 101–110 (1985).
25. Okamoto, H. et al. *EMBO J.* **5**, 589–596 (1986).
26. Rubin, G. M. & Spradling, A. C. *Science* **218**, 348–353 (1982).
27. O'Donnell, P. T. & Bernstein, S. I. *J. Cell Biol.* **107**, 2601–2612 (1988).

ACKNOWLEDGEMENTS. We thank Meg Stark for help with electron microscopy. This work was supported by SERC.

Non-vesicular release of glutamate from glial cells by reversed electrogenic glutamate uptake

Marek Szatkowski, Boris Barbour & David Attwell

Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK

GLUTAMATE uptake into nerve and glial cells usually functions to keep the extracellular glutamate concentration low in the central nervous system¹. But one component of glutamate release from neurons is calcium-independent, suggesting a non-vesicular release that may be due to a reversal of glutamate uptake^{2,3}. The activity of the electrogenic glutamate uptake carrier can be monitored by measuring the membrane current it produces⁴, and uptake is activated by intracellular potassium ions⁵. Here we report that raising the potassium concentration around glial cells evokes an outward current component produced by reversed glutamate uptake. This current is activated by intracellular glutamate and sodium, inhibited by extracellular glutamate and sodium, and increased by membrane depolarization. These results demonstrate a non-vesicular mechanism for the release of glutamate from glial cells and neurons. This mechanism may contribute to the neurotoxic rise in extracellular glutamate concentration during brain anoxia.

Electrogenic glutamate transport was studied in whole-cell clamped Müller cells isolated from the salamander retina⁴. The glutamate uptake system in these cells has properties similar to that in mammalian glia^{6,7} and neurons⁸. Normal uptake of glutamate is activated by intracellular potassium (probably because the uptake carrier transports K^+ out of the cell as well as transporting Na^+ in) and generates an inward membrane current⁵. We tested, therefore, whether extracellular potassium activates an outward membrane current (reflecting glutamate release) in cells containing sodium and glutamate. To minimize

contamination from other potassium-dependent membrane currents: the sodium-potassium exchange pump was inhibited with 100 μ M ouabain⁹; the inward rectifier channels that contribute much of the cells' potassium conductance were largely blocked with 6 mM barium¹⁰; in most experiments potassium was omitted from the patch pipette to reduce greatly the cells' potassium conductance; and we selected cells that had lost their vitreal endfeet in the cell isolation procedure as the endfeet contain over 90% of the cells' inward rectifier channels^{11,12}.

In cells studied with patch pipettes containing at least 1 mM sodium and 30 μ M glutamate, raising the extracellular potassium concentration from 0 to 10 mM evoked an outward membrane current (Fig. 1a). This current was largely blocked in the presence of 100 μ M or 1 mM extracellular glutamate (Fig. 1b), and was stimulated three- to fourfold when external sodium ions were replaced by choline (Fig. 1b). It was absent if glutamate or sodium were omitted from the pipette (see below). These results are consistent with the idea that the outward membrane current reflects outward transport of glutamate (with an excess of sodium ions) by the uptake carrier, and that a raised $[Na]_o$ or $[glutamate]_o$ hinders the loss of sodium or glutamate from the carrier at the outer membrane surface. The magnitude of the current was larger at positive potentials (Fig. 1a, c).

To confirm that the current changes reflect the activity of the glutamate uptake carrier, we investigated the pharmacology of the suppressive action of external excitatory amino acids on the potassium-evoked current. Consistent with the known pharmacology of the uptake carrier, the glutamate analogues that are relatively specific for acting on ion channels (*N*-methyl-D-aspartate (NMDA), quisqualate and kainate) produced a much smaller suppression than did L-glutamate or L- or D-aspartate (Fig. 2a, b). On average, different glutamate analogues (100 μ M) suppressed the current evoked by 10 mM extracellular potassium by the following percentages ($\pm$ s.e.): D-aspartate $95 \pm 2\%$; L-aspartate $94 \pm 2\%$; L-glutamate $90 \pm 2\%$; NMDA $31 \pm 2\%$; quisqualate $28 \pm 4\%$; kainate $12 \pm 3\%$; D-glutamate $10 \pm 3\%$ (5 cells studied at -5 mV for each analogue and 15 cells for L-glutamate).

FIG. 1 External potassium activates reversed glutamate uptake. **a**, Current changes in a Müller cell voltage-clamped to various membrane potentials (shown by each trace) when $[K]_o$ was raised from 0 to 10 mM (timing shown by black bars). Potassium evokes an outward current, larger at positive potentials, reflecting glutamate release from the cell. **b**, Inhibition of reversed glutamate uptake by external sodium and glutamate (at -5 mV). During repeated increases of $[K]_o$ from 0 to 10 mM (black bars), the external solution was changed from normal (barium) Ringer's to Ringer's containing 1 mM glutamate (which inhibited the reversed uptake current), and then to sodium-free Ringer's (which increased the reversed uptake current). In the presence of external glutamate (or if sodium or glutamate were omitted from the patch pipette), 10 mM $[K]_o$ sometimes evoked a small inward current (~ 2 pA), which we attribute to residual current flow through potassium channels; this was corrected for when quantitatively analysing the K^+ -evoked currents in the absence of external glutamate. **c**, Voltage-dependence of the reversed uptake current evoked (in a different cell) by 10 mM $[K]_o$ in normal Ringer's, Ringer's containing 1 mM glutamate and in sodium-free Ringer's. In Ringer's with glutamate, little K^+ -evoked current was seen at any potential.

METHODS. Glial (Müller) cells isolated from the retina of the tiger salamander⁴ were whole-cell clamped (series resistance ~ 2 M Ω) using pipettes containing the 100 mM sodium glutamate solution described in Fig. 3c legend ($x=100$). External solution normally contained (in mM): NaCl, 100; KCl, $x=0$ or 10; choline chloride, $10-x$; $MgCl_2$, 0.5; $CaCl_2$, 3; HEPES buffer, 5; glucose, 15; $BaCl_2$, 6; ouabain, 0.1; pH adjusted to 7.35 with 2 mM *N*-methyl-D-glucamine (NMDG). For sodium-free solution, NaCl was replaced by choline chloride.

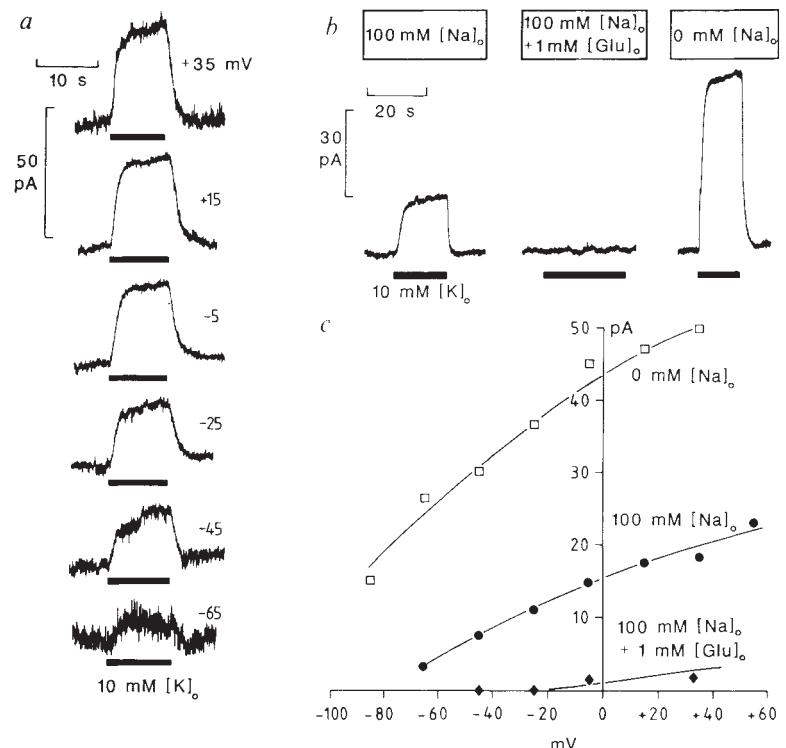
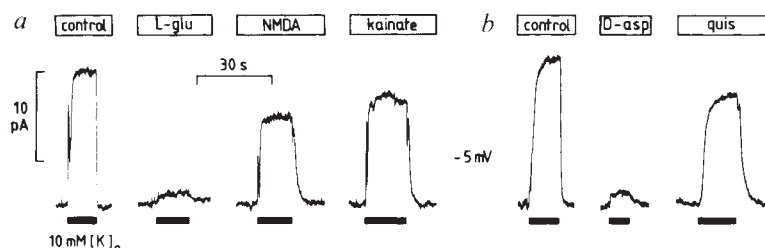


FIG. 2 Pharmacology of glutamate's action on reversed uptake. Pipette solutions contained 100 mM sodium-glutamate as in Fig. 3c ($x=100$). External (sodium-containing) solution was as in Fig. 1. *a* and *b* Effect of glutamate analogues (100 μ M) on the reversed uptake current evoked at -5 mV by a rise in $[K]_o$ from 0 to 10 mM (black bars). In one cell (*a*) in the absence of glutamate or analogues 10 mM $[K]_o$ evokes an outward current, which is largely inhibited by L-glutamate, but much less so by NMDA or kainate. In another cell (*b*) the control current is greatly inhibited by D-aspartate, but much less so by quisqualate (quis). As well as suppressing the potassium-evoked current, glutamate also altered the current recorded in the absence of potassium, and the pharmacology of this current change (assessed with the analogues used in Fig. 2*a* and *b*) was appropriate for glutamate acting on the uptake carrier rather than on ion channels²². The glutamate-evoked current was outward at potentials above, and inward at potentials below -40 mV. Omitting glutamate and sodium from the pipette abolished the outward current and greatly reduced the inward current, again

In addition to suppressing the potassium-evoked current, glutamate also changed the baseline current recorded in the absence of potassium. We tentatively attribute this current change to a suppression of two potassium-independent modes of reversed glutamate uptake (see Fig. 2 legend); these were not studied in detail.

The dependence of potassium-evoked reversed uptake on pipette (and hence intracellular) glutamate concentration (Fig. 3*a, b*) could be roughly fitted by a Michaelis-Menten equation with a Michaelis constant K_m of 450 μ M. The dependence of reversed uptake on pipette sodium concentration (Fig. 3*c, d*) could be described by a Hill equation with a Hill coefficient of



indicating that these currents are not generated by glutamate-gated ion channels. The outward current may reflect suppression of an efflux of glutamate anions on the uptake carrier, unassociated with the net transport of other ions. The inward current may reflect suppression of an efflux of glutamate cotransported with an excess of Na^+ ions. Typically, the glutamate-evoked outward and inward currents at -5 and -65 mV were about 35% and 43% the size of the current evoked at -5 mV by 10 mM $[K]_o$.

2, with half-activation at 3 mM $[Na]_i$. These results are consistent with one glutamate anion and at least two sodium ions being transported out of the cell on each carrier cycle. They suggest that in the physiological situation (where $[Na]_i$ is about 25 mM¹³ and $[glutamate]_i$ is about 10 mM¹⁴⁻¹⁶) reversed uptake will not be rate-limited by sodium- or glutamate-binding at the inner membrane surface.

The dependence of the potassium-evoked current on $[K]_o$ could be fitted by a Michaelis-Menten relation with a K_m of 19 mM (Fig. 4*a, b*). The stimulatory effect of extracellular potassium on reversed glutamate uptake is the opposite of its inhibitory action on normal uptake⁵. This strongly supports the

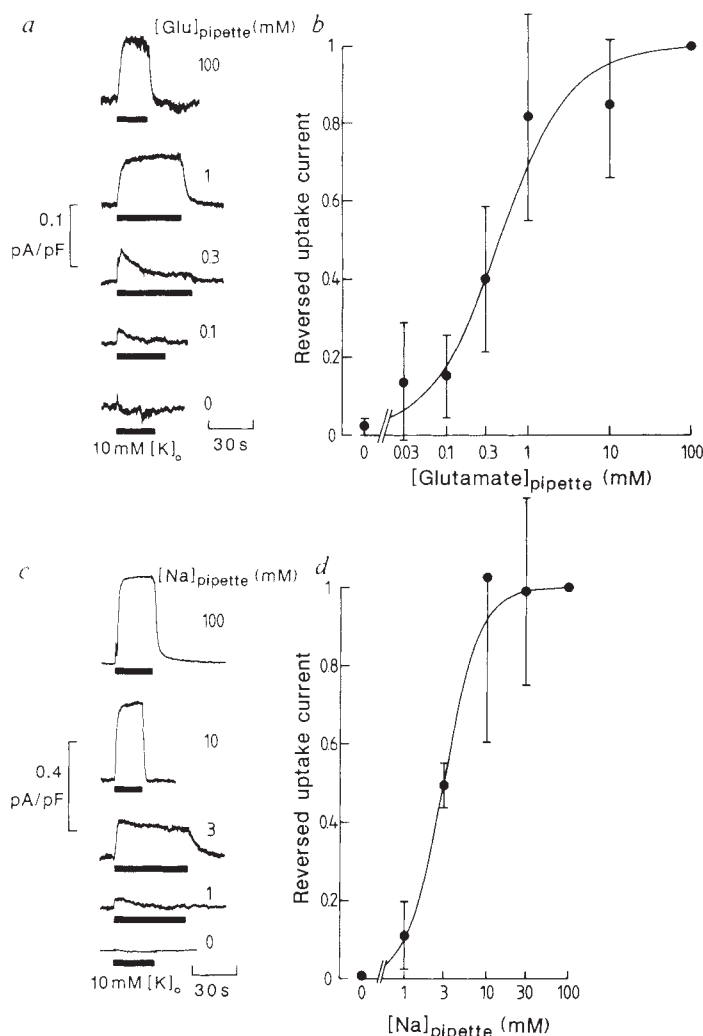


FIG. 3 Dependence of K^+ -dependent reversed glutamate uptake on the intracellular glutamate and sodium concentrations. *a*, Responses to a rise in $[K]_o$ (0 to 10 mM) of five cells clamped to 0 mV with pipettes containing different L-glutamate concentrations. Pipettes contained (in mM): sodium-glutamate, x ($x=0$ to 100); NaCl, 100- x ; (NMDG)₂-EGTA, 5; Mg-ATP, 5; HEPES buffer, 5; CaCl₂, 1; MgCl₂, 2; sodium malonate, 0.2; sodium oxalomalate (OM), 2; aminooxyacetic acid (AOAA), 5; methionine sulfoximine (MSO), 2; pH adjusted to 7.0 with 26 mM NMDG. Malonate and OM, AOAA and MSO were included to block the Krebs cycle, glutamate transaminase and glutamine synthetase respectively, to provide better control of the intracellular glutamate concentration by the patch pipette. External solution (sodium-containing) was as in Fig. 1. Data were normalized by cell capacitance to reduce variation caused by variation in cell size⁵. When $[glutamate]_{pipette}$ (or $[Na]_{pipette}$ as in *c*) were low, responses to raised $[K]_o$ showed a sag after an initial peak. We attribute this to depletion of intracellular glutamate or sodium: for 10 pA reversed carrier current and a cell volume of $5 \times 10^{-15} m^3$ $[Glu]_i$ and $[Na]_i$ will be depleted by $20 \mu M s^{-1}$ and $60 \mu M s^{-1}$ respectively (assuming three Na^+ and one K^+ are co- and counter-transported with glutamate, and ignoring replenishment from the patch pipette). For quantitative analysis initial peak currents were measured. *b*, Mean currents ($\pm$ s.d.) from experiments as in *a* for five cells studied at each glutamate concentration (normalized to data from 5 cells studied with 100 mM $[glutamate]$ on the same day: mean current with 100 mM $[Glu]$ was 0.1 pA/pF). Curve is $[Glu]/([Glu] + K_m)$, with $K_m=450 \mu M$ (best fit Hill coefficient was 0.71). *c*, Responses (at 0 mV) of five cells with different internal sodium concentrations to a rise in $[K]_o$. Pipettes contained (mM): sodium-glutamate, x ($x=0$ to 100); NMDG-glutamate, 100- x ; (NMDG)₂-EGTA, 5; Mg-ATP, 5; HEPES buffer, 5; CaCl₂, 1; MgCl₂, 2; pH adjusted to 7.0 with 14 mM NMDG. External solution was sodium-free barium Ringer's plus ouabain as in Fig. 1*b*: sodium-free solution facilitated attainment of low $[Na]_i$ in the cell when $[Na]_{pipette}$ was low. *d*, Mean currents ($\pm$ s.d.) from experiments as in *c* for 4-6 cells at each sodium concentration (normalized to data from 4-6 cells studied with 100 mM $[Na]$ on the same day). Curve is $[Na]^2/([Na]^2 + K^2)$, with $K=3$ mM (best-fit Hill coefficient was 1.89).

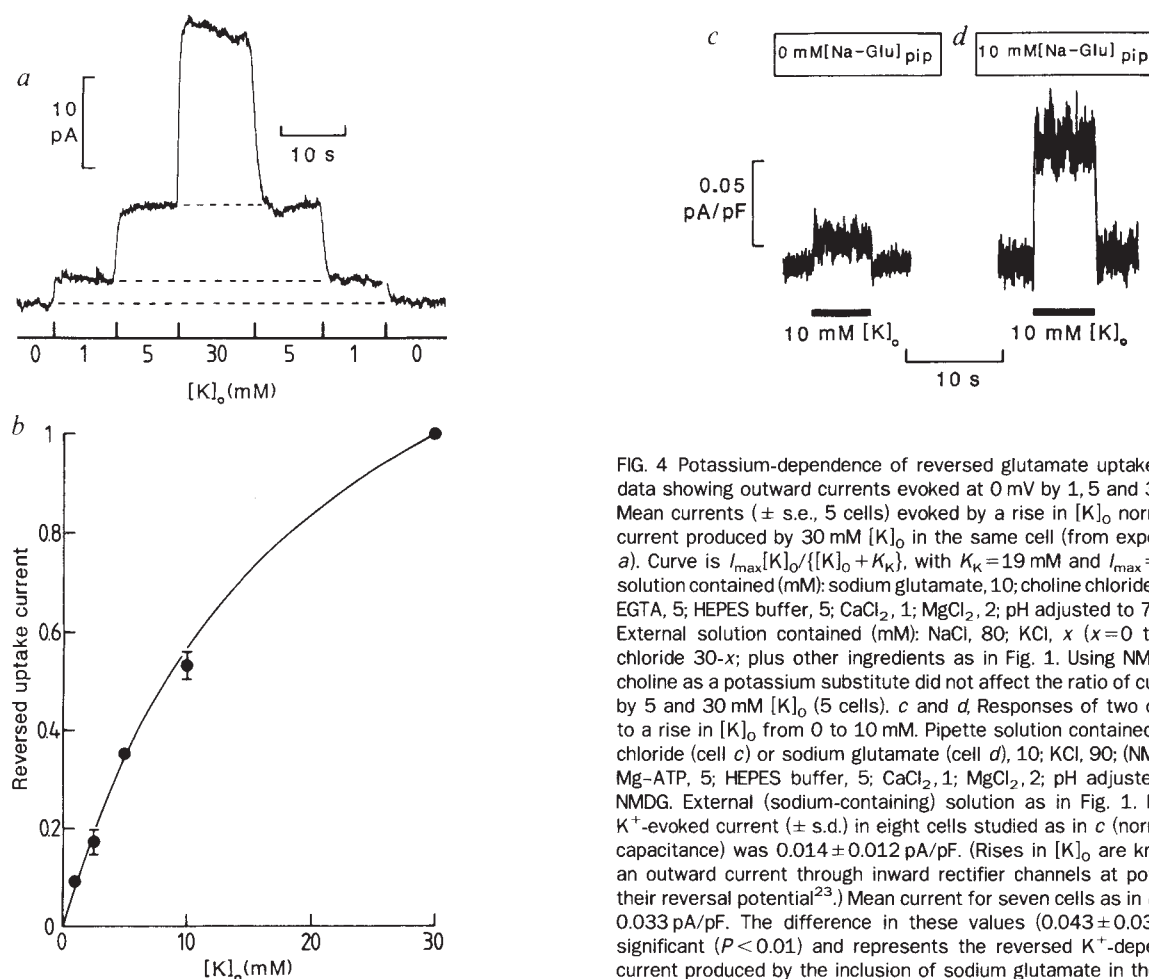


FIG. 4 Potassium-dependence of reversed glutamate uptake. *a*, Specimen data showing outward currents evoked at 0 mV by 1, 5 and 30 mM $[K]_o$. *b*, Mean currents ($\pm$ s.e., 5 cells) evoked by a rise in $[K]_o$ in the same cell (from experiments as in *a*). Curve is $I_{max}[K]_o/([K]_o + K_K)$, with $K_K = 19$ mM and $I_{max} = 1.62$. Pipette solution contained (mM): sodium glutamate, 10; choline chloride, 92; (NMDG) $_2$ -EGTA, 5; HEPES buffer, 5; CaCl $_2$, 1; MgCl $_2$, 2; pH adjusted to 7.0 with NMDG. External solution contained (mM): NaCl, 80; KCl, x ($x = 0$ to 30); choline chloride 30- x ; plus other ingredients as in Fig. 1. Using NMDG instead of choline as a potassium substitute did not affect the ratio of currents evoked by 5 and 30 mM $[K]_o$ (5 cells). *c* and *d*, Responses of two cells (at 0 mV) to a rise in $[K]_o$ from 0 to 10 mM. Pipette solution contained (mM): choline chloride (cell *c*) or sodium glutamate (cell *d*), 10; KCl, 90; (NMDG) $_2$ -EGTA 5; Mg-ATP, 5; HEPES buffer, 5; CaCl $_2$, 1; MgCl $_2$, 2; pH adjusted to 7.0 with NMDG. External (sodium-containing) solution as in Fig. 1. Mean outward K^+ -evoked current ($\pm$ s.d.) in eight cells studied as in *c* (normalized to cell capacitance) was 0.014 ± 0.012 pA/pF. (Rises in $[K]_o$ are known to evoke an outward current through inward rectifier channels at potentials above their reversal potential²³.) Mean current for seven cells as in *d* was 0.057 ± 0.033 pA/pF. The difference in these values (0.043 ± 0.035 pA/pF) was significant ($P < 0.01$) and represents the reversed K^+ -dependent uptake current produced by the inclusion of sodium glutamate in the pipette.

notion that K^+ ions can be transported on the uptake carrier (rather than extracellular K^+ altering the rate of uptake by an allosteric mechanism). From the results above we propose, by analogy with the stoichiometry of normal K^+ -dependent glutamate uptake⁵, that potassium-dependent reversed uptake involves the movement of one potassium ion into the cell on each carrier cycle, together with the movement of one glutamate anion and three Na^+ (or two Na^+ and one H^+) ions out of the cell.

The experiments described above were carried out with no potassium in the patch pipette. To examine the magnitude of reversed K^+ -dependent uptake in the more physiological situation with potassium in the cell, we compared the effect of raising $[K]_o$ on cells clamped with pipettes containing 90 mM KCl and either 10 mM sodium glutamate or 10 mM choline chloride. With no sodium glutamate in the pipette, raising $[K]_o$ from 0 to 10 mM produced a small outward current at 0 mV, probably through those inward rectifier channels that remain unblocked by 6 mM barium (Fig. 4c). With 10 mM sodium glutamate in the pipette, the current was larger (Fig. 4d), presumably because of the extra outward current generated by reversed glutamate uptake. The magnitude of this extra current (0.043 ± 0.035 (s.d.) pA/pF, see Fig. 4 legend) was approximately half that of the current produced by 10 mM $[K]_o$ in cells studied without K^+ in the pipette, consistent with a high $[K]_i$ hindering the loss of K^+ from the carrier at the inner membrane surface.

We have shown previously that a rise in $[K]_o$ inhibits glutamate uptake⁵. The data in Fig. 1 and Fig. 4a suggest that, when $[glutamate]_o$ is low, membrane depolarization or a rise in extracellular potassium concentration will stimulate the release of glutamate (and aspartate) by reversed potassium-dependent glutamate uptake. As neuronal glutamate uptake has properties

similar to that in glial cells^{5,8}, this mechanism could underlie the calcium-independent release of glutamate from retinal photoreceptors² and from cerebral cortical synaptosomes³.

During brain anoxia the extracellular glutamate concentration rises to neurotoxic levels¹⁷. The origin of this extracellular glutamate is uncertain, however, because vesicular release of glutamate is known to be inhibited when ATP levels fall after a few minutes anoxia¹⁸. During anoxia the ion and voltage gradients favouring glutamate uptake are greatly reduced (the extracellular potassium concentration can rise to 50 mM, depolarizing glia to -20 mV, and $[Na]_o$ falls to 60 mM^{19,20}). In this situation reversed uptake will release glutamate into the extracellular space, contributing to the rise in extracellular glutamate concentration. The expression for the reversal potential of K^+ -dependent glutamate transport⁵ predicts that (at 37 °C, with $[Glu]_i = 10$ mM, $V_{Na} = 24$ mV and $V_K = V_m = -20$ mV) the carrier could release glutamate until the extracellular glutamate concentration reaches either 150 μ M (assuming a 3 Na^+ : 1 Glu^- : 1 K^+ stoichiometry) or 370 μ M (assuming 2 Na^+ and 1 H^+ are cotransported, 1 K^+ is counter-transported and $V_H = 0$ mV during anoxia²⁰). More of the intracellular glutamate in the brain may be available for release by reversed uptake than by vesicular release^{18,21}. It is possible, therefore, that reversed uptake has an important role in causing neuronal death in conditions such as stroke and perinatal asphyxia. □

Received 20 July; accepted 1 October 1990.

1. Hertz, L. *Prog. Neurobiol.* **13**, 277–323 (1979).
2. Miller, A. M. & Schwartz, E. A. *J. Physiol., Lond.* **334**, 325–350 (1983).
3. Nicholls, D. G., Sihra, T. S. & Sanchez-Prieto, J. *J. Neurochem.* **49**, 50–57 (1987).
4. Brew, H. & Attwell, D. *Nature* **327**, 707–709 (1987).
5. Barbour, B., Brew, H. & Attwell, D. *Nature* **335**, 433–435 (1988).
6. Sarantis, M. & Attwell, D. *Brain Res.* **516**, 322–325 (1990).

7. Cull-Candy, S. G., Howe, J. R. & Ogden, D. *J. Physiol., Lond.* **400**, 189–222 (1988).
8. Kanner, B. I. & Schuldiner, S. *Crit. Rev. Biochem.* **22**, 1–38 (1987).
9. Walz, W. & Hertz, L. *J. Neurochem.* **39**, 70–77 (1982).
10. Newman, E. A. *Nature* **317**, 809–811 (1985).
11. Newman, E. A. *Nature* **309**, 155–157 (1984).
12. Brew, H., Gray, P., Mobbs, P. & Attwell, D. *Nature* **324**, 466–468 (1986).
13. Ballanyi, K., Grafe, P. & Ten Bruggencate, G. *J. Physiol., Lond.* **382**, 159–174 (1987).
14. Berger, S. J., Carter, J. G. & Lowry, O. H. *J. Neurochem.* **28**, 149–158 (1977).
15. Schousboe, A., Fosmark, H. & Hertz, L. *J. Neurochem.* **25**, 909–911 (1975).
16. Kvamme, E., Schousboe, A., Hertz, L., Torgner, I. A. & Svenneby, G. *Neurochem. Res.* **10**, 993–1008 (1985).
17. Rothman, S. M. & Olney, J. W. *Trends Neurosci.* **10**, 299–302 (1987).
18. Sanchez-Prieto, J. & Gonzalez, P. *J. Neurochem.* **50**, 1322–1324 (1988).
19. Walz, W. & Hertz, L. *Prog. Neurobiol.* **20**, 133–183 (1983).
20. Siesjö, B. *News physiol. Sci.* **5**, 120–125 (1990).
21. Ikeda, M., Nakazawa, T., Abe, K., Kaneko, T. & Yamatsu, K. *Neurosci. Lett.* **96**, 202–206 (1989).
22. Sontheimer, H., Kettenmann, H., Backus, K. H. & Schachner, M. *Glia* **1**, 328–336 (1988).
23. Sakmann, B. & Trube, G. *J. Physiol. Lond.* **347**, 641–657 (1984).

ACKNOWLEDGEMENTS. Supported by the Wellcome Trust, M.R.C. and S.E.R.C.

A nucleoprotein peptide of influenza A virus stimulates assembly of HLA-B27 class I heavy chains and β_2 -microglobulin translated *in vitro*

Sune Kvist & Ute Hamann*

Ludwig Institute for Cancer Research, Stockholm Branch, Box 60202, S-10401 Stockholm, Sweden

MOST cytotoxic T lymphocytes (CTL) recognize epitopes of foreign viral proteins in association with class I major histocompatibility complex (MHC) molecules¹. Viral proteins synthesized in the cytoplasm require intracellular fragmentation and exposure to the class I antigens for the development of CTL responses^{2–4}. Although indirect evidence for binding of peptides to class I antigens has accumulated, direct binding has only been shown recently^{5,6}. The formation of complexes between peptide and class I antigen may occur in the endoplasmic reticulum (ER)^{7,8} and peptides have been shown to induce assembly of the class I complex⁹. We have translated the messenger RNAs encoding HLA-B27 (subtype 2705)^{10,11} and β_2 -microglobulin in a rabbit reticulocyte lysate supplemented with human microsomal membranes (to mimic ER membranes), in the absence and presence of a peptide derived from the nucleoprotein¹² (residues 384–394) of influenza A virus. This peptide induces CTL activity against target cells expressing the HLA-B27 antigen¹³. Here we report direct evidence that the nucleoprotein peptide promotes assembly of the HLA-B27 heavy chain and β_2 -microglobulin, and that this can occur in the ER immediately after synthesis of the two proteins.

The complementary DNA clones encoding the HLA-B27 heavy chain and β_2 -microglobulin (β_2m) were inserted into the vector pGem-3Zf(–) and mRNA was transcribed by SP6 RNA polymerase¹⁴. The mRNAs were separately translated *in vitro* in a lysate prepared from rabbit reticulocytes and supplemented with purified microsomes from the human lymphoid cell line Raji. Immunoprecipitations of [³⁵S]methionine-labelled HLA-B27 heavy chain and β_2m were analysed by SDS-PAGE. The HLA-B27 heavy chain appears as a doublet (relative molecular mass ~39,000 (M_r 39K)) whereas β_2m migrates as one distinct band (M_r 12K; Fig. 1, lanes 1 and 2, respectively). However, co-translation of both mRNAs and immunoprecipitation with a monoclonal antibody (W6/32) (ref. 15) recognizing the assembled form exclusively, show that only a small fraction (<5%) of the translated material is present in an assembled

form (Fig. 1, lane 3), although translation efficiency of both chains is unchanged (data not shown).

To characterize our system further, we examined the microsomal membranes for their endogenous HLA heavy chain and β_2m content. This also allowed us to determine whether the radiolabelled chains contributed to the overall chemical amounts of HLA heavy and light chains within the microsomes. Increasing amounts of mRNA for either HLA-B27 heavy chain or β_2m were translated, followed by analyses of both the radiochemical amounts and the total amounts of HLA heavy and light chains.



FIG. 1 Cell-free translation and SDS-PAGE analysis of HLA-B27 and β_2m . Lane 1, translation of HLA-B27 mRNA and immunoprecipitation with a rabbit anti-HLA serum specific for the heavy chain. Lane 2, translation of β_2m mRNA and immunoprecipitation with the monoclonal antibody BBM.1 (ref. 20) (anti- β_2m). Lane 3, co-translation of the mRNAs for HLA-B27 and β_2m and immunoprecipitation with the monoclonal W6/32 (ref. 15), specific for the assembled form of the HLA antigens. The lower band in the HLA-B27 doublet is presumably the non-glycosylated form.

METHODS. The mRNA was transcribed from cDNA clones of HLA-B27 and β_2m , inserted into the EcoRI site of pGem-3Zf(–), using SP6 RNA polymerase (Promega)²¹. A modified translation-enhancing fragment of the E3/19K gene of adenovirus type 2 was inserted upstream of the AUG initiation codon for the HLA-B27 cDNA. Rabbit reticulocyte lysate was prepared essentially as described previously^{22,23}. During translation of HLA-B27 mRNA (50 ng) and β_2m (150 ng) the lysate was supplemented with 3 μ l of purified microsomes (60 A₂₈₀ ml^{–1}) (ref. 24) from the Raji lymphoid cell line. Reaction mixture (65 μ l total) was: lysate 46 μ l, mRNA 2–4 μ l, [³⁵S]methionine 6 μ l (90 μ Ci, >1,000 Ci mmol^{–1}, Amersham), microsomal membranes 3 μ l, RNasin 10 units (Promega), phenylmethylsulphonyl fluoride (PMSF) 2 μ g and aprotinin 10 units. Translation was started by addition of the mRNA and incubation at 37 °C. Translation was terminated after 90 min by changing the temperature to 0 °C and adding 135 μ l of TNE buffer (20 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA). The membranes were pelleted in an Eppendorf centrifuge for 5 min and solubilized in a physiological buffer containing 5 mM EDTA and 1% Triton-X100. Non-solubilized material was pelleted as above and the supernatant was immunoprecipitated as described elsewhere²⁵. SDS-PAGE analysis as described previously²⁶. The rabbit antiserum against HLA heavy chains was prepared by injecting purified HLA antigens. The antiserum was absorbed on β_2m coupled to Sepharose beads and does not react with β_2m .

* Present address: German Institute for Cancer Research, Department for Experimental Pathology, Im Neuenheimer Feld 280, W-6900, Heidelberg, Germany.

After translation, each lysate was divided into two equal parts and the microsomes were pelleted and washed before solubilization. One part was immunoprecipitated with a rabbit antiserum directed against HLA heavy chains, and was analysed by SDS-PAGE. The other part, containing the entire pool of HLA antigens present in the microsomes, was analysed immediately by SDS-PAGE and western blotting.

We used amounts of HLA-B27 heavy chain mRNA ranging from 0 to >500 ng; the HLA-B27 band becomes visible at ~32 ng (Fig. 2a, lanes 1–5). Despite the efficient translation and large amount of radiochemical HLA-B27 heavy chain (Fig. 2a, lane 5), there was no detectable increase in the total amount of HLA heavy chains present in the microsomes (Fig. 2a, lanes 6–10). Thus, the amount of HLA-B27 heavy chain translated *in vitro* in the reticulocyte lysate does not contribute significantly to the large pool of endogenous HLA antigen heavy chains in the microsomes. We analysed β 2m in the same way and found that, similarly, there is a large excess of endogenous β 2m in the microsomes and that the amount translated *in vitro* does not give rise to an increase detectable by western blotting (Fig. 2b; compare lanes 1–5 with 6–10). Thus, endogenous microsomal HLA, both heavy and light chains, is present in large excess over newly translated radiolabelled HLA-B27 heavy chain and β 2m.

We tested whether addition of a peptide¹³ from the nucleoprotein (NP) of influenza A virus, known to constitute a target for CTL in a HLA-B27 restricted manner, would promote assembly. Messenger RNAs for both HLA-B27 heavy chain and β 2m were cotranslated with increasing amounts of the NP 384–394 peptide. In the absence of peptide, only small amounts of assembled HLA-B27 heavy chain and β 2m are found (Fig. 3, lane 1). At a peptide concentration of 1 μ M, little increase in assembly is observed (Fig. 3, lane 2), but at 4 μ M (lane 3) enhanced assembly is seen, which becomes more apparent at 16 μ M (lane 4). The increase in assembly was observed only as enhanced intensity in the band for the heavy chain of HLA-B27.

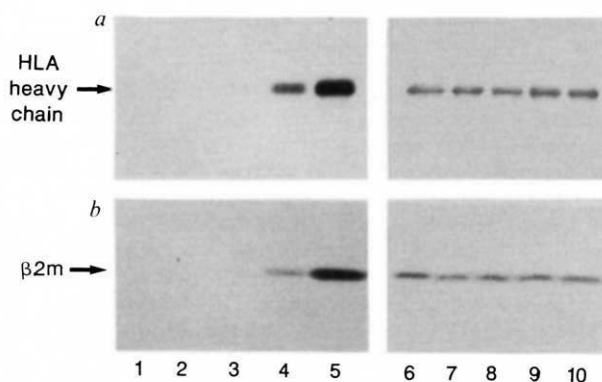


FIG. 2 Analysis of HLA antigen content in Raji microsomal vesicles. *a*, Increasing amounts of mRNA for HLA-B27 were translated *in vitro* in the presence of Raji microsomes (3 μ l, A_{280} 60 units ml^{-1}). Lane 1, absence of mRNA; lane 2, 8 ng of mRNA; lane 3, 32 ng of mRNA; lane 4, 128 ng of mRNA; lane 5, 512 ng of mRNA. Lanes 6–10, the same amounts of microsomes as used in lanes 1–5 were analysed by western blotting using an excess of a rabbit antiserum against HLA heavy chains. *b*, Increasing amounts of mRNA for β 2m were translated at the same concentrations as in *a* (lanes 1–5). Lanes 6–10, the same amounts of microsomes as used in lanes 1–5 were analysed by western blotting using a rabbit antiserum against β 2m. METHODS. Translation was as described in the legend to Fig. 1, but with increasing amounts of mRNA. After translation, the samples were solubilized and each separate translation was divided into two equal parts. One part was used for immunoprecipitation with the rabbit antiserum against the HLA heavy chain (*a*), or a rabbit antiserum against human β 2m (*b*). The second part was used for western blotting with the same antisera, respectively. The rabbit antiserum against human β 2m was prepared in this laboratory.

The intensity of the band corresponding to β 2m was unaffected by changes in the concentration of the NP 384–394 peptide. This specificity was further demonstrated using a nucleoprotein peptide specific for the mouse MHC class I K^d antigen¹⁶. This NP 147–158 peptide did not induce assembly when used at the same concentrations as the NP 384–394 peptide (Fig. 3, lanes 5–7). The band for β 2m showed the same intensity as when the HLA-B27 peptide was used. We conclude that assembly of HLA-B27 and β 2m was specifically induced or promoted by the NP 384–394 peptide.

We suggest that newly translated radiolabelled β 2m complexes with the large excess of endogenous unlabelled HLA heavy chains in the microsomal vesicles. The radiolabelled HLA-B27 heavy chain probably assembles mostly with microsomal unlabelled β 2m. Only a small fraction of newly translated radiolabelled β 2m associates with HLA-B27 heavy chain and does not visibly contribute to the labelled β 2m band. This explanation for the constant β 2m band shown in Fig. 3 is supported both by the finding shown in Fig. 2, and the fact that the amount of radiolabelled β 2m precipitated in the experiment in Fig. 3 constitutes a very small fraction of the labelled β 2m (data not shown). Thus, it should also be possible to observe enhanced assembly of HLA-B27 heavy chains with β 2m without *in vitro* translation of β 2m mRNA.

To test this, we translated the HLA-B27 heavy chain mRNA *in vitro*, and immunoprecipitated the radiolabelled proteins with two different monoclonal antibodies directed against the assembled HLA complexes (W6/32) and against β 2m (BBM.1), respectively. The W6/32 antibody precipitates increasing

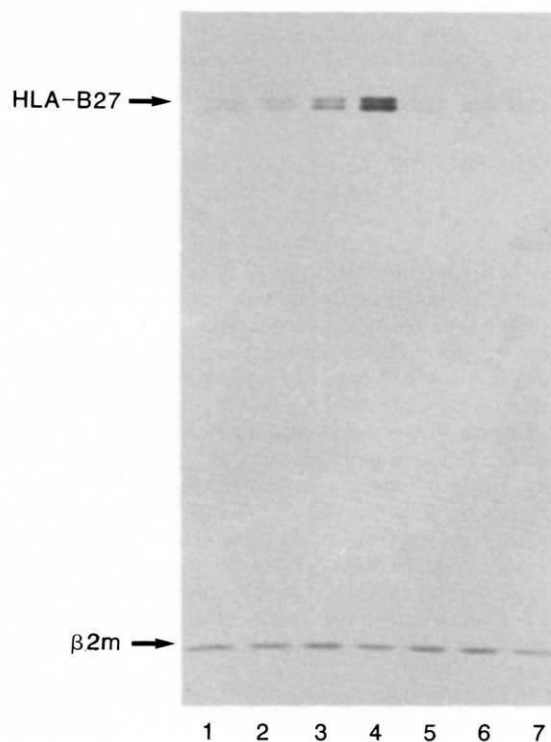


FIG. 3 Stimulation of HLA-B27 antigen and β 2m assembly by the NP 384–394 peptide. Messenger RNA for HLA-B27 and β 2m were translated *in vitro*, and immunoprecipitations were carried out with the monoclonal antibody W6/32. Lane 1, absence of peptide. Lanes 2–4, NP 384–394 peptide concentrations of 1.0, 4.0 and 16 μ M, respectively. Lanes 5–7, concentrations of NP 147–158 peptide¹⁶ (H-2 K^d -specific) were 1.0, 4.0 and 16 μ M, respectively.

METHODS. *In vitro* translation was carried out as described in the legend to Fig. 1, but the peptides were added before incubation at 37 °C at the concentrations stated above. Peptides were dissolved in water and were: NP 384–394 (ref. 13), Arg-Tyr-Trp-Ala-Ile-Arg-Thr-Arg-Ser-Gly-Gly; NP 147–158 (refs 16, 17), Thr-Tyr-Gln-Arg-Thr-Arg-Ala-Leu-Val-Arg-Thr-Gly.

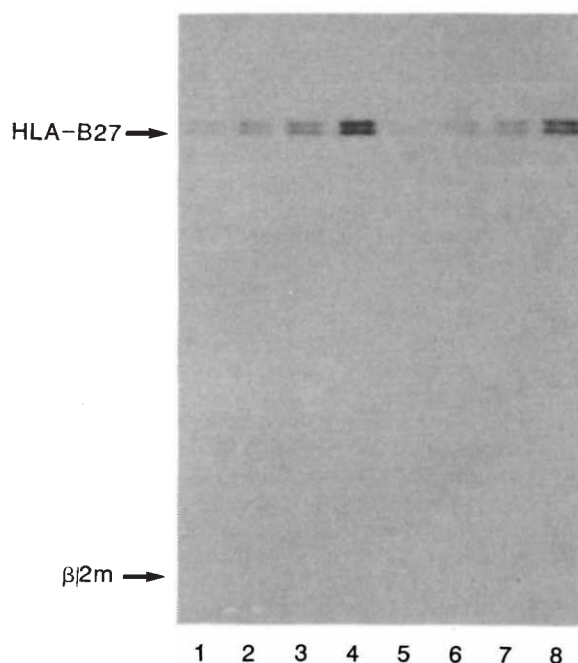


FIG. 4 *In vitro* translated HLA-B27 heavy chain assemblies with endogenous microsomal $\beta 2m$. Messenger RNA (50 ng) for HLA-B27 heavy chain was translated in the presence of Raji microsomes (3 μ l) at different concentrations of NP 384–394 peptide; lanes 1 and 5, absence of peptide; lane 2 and 6, 1 μ M; lanes 3 and 7, 4 μ M; lanes 4 and 8, 16 μ M. Immunoprecipitations were carried out with monoclonal antibody W6/32 (lanes 1–4), and BBM.1 (lanes 5–8). *In vitro* translations were as described in the legend to Fig. 1.

amounts of HLA-B27 heavy chains when the concentration of NP 384–394 is increased (Fig. 4, lanes 1–4). Thus, we observe enhanced assembly, even without translation of $\beta 2m$ mRNA. The antibody BBM.1 confirms this result (Fig. 4, lanes 5–8); it is possible to immunoprecipitate the HLA-B27 heavy chain, and to observe enhanced assembly with peptide NP 384–394, with a monoclonal antibody directed against $\beta 2m$.

Here, we have shown that a small peptide can stimulate assembly of HLA-B27 class I heavy chain with $\beta 2m$. This occurs in the lumen of microsomal vesicles that mimic the ER compartment. Thus, it seems that assembly of MHC class I molecules can occur in this compartment. The increase of assembly is roughly 3–8-fold (when analysed by densitometry) over that observed in the absence of the specific peptide. Overexposure of the x-ray films indicates that some HLA-B27– $\beta 2m$ assembly occurs in the absence of the NP 384–394 peptide. This is probably due to stimulation either by endogenous peptides, or spontaneous association of the chains without peptides involved. Such 'empty' MHC class I antigens have recently been demonstrated^{17,18}.

Our results confirm the findings of Townsend and colleagues, who found that assembly and cell surface expression of murine H-2 antigens can be stimulated by specific peptides in the cell line RMA-S (ref. 9). This implies that our system, supplemented with human microsomes, probably resembles the *in vivo* assembly process. Assembly of MHC class I antigens can occur in the lysate of detergent-solubilized cells, and this process is dependent on both peptide and $\beta 2m$ concentrations¹⁹. In our system we were unable to detect assembly after solubilization (data not shown). The peptide concentrations used in this study are similar to those used by Townsend *et al.*¹⁹. As the different components of our system can easily be varied, it will be useful in elucidating the complicated nature of MHC class I antigen assembly. □

Received 30 July; accepted 26 September 1990.

1. Zinkernagel, R. M. & Doherty, P. C. *Adv. Immun.* **27**, 51–77 (1979).
2. Townsend, A. R. M., Gotch, F. M. & Davey, J. *Cell* **42**, 457–467 (1985).
3. Townsend, A. R. M. *et al.* *Cell* **44**, 959–968 (1986).
4. Maryanski, J. L., Pala, P., Corradin, G., Jordan, B. R. & Cerottini, J. C. *Nature* **324**, 578–579 (1986).
5. Chen, B. P. & Parham, P. *Nature* **337**, 743–745 (1989).
6. Bouillot, M. *et al.* *Nature* **339**, 473–475 (1989).
7. Nuchtern, J. G., Bonifacio, J. S., Biddison, W. E. & Klausner, R. D. *Nature* **339**, 223–226 (1989).
8. Yewdell, J. W. & Bennink, J. R. *Science* **244**, 1072–1075 (1989).
9. Townsend, A. R. M. *et al.* *Nature* **340**, 443–448 (1989).
10. Szöts, H., Riethmüller, G., Weiss, E. & Meo, T. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1428–1432 (1986).
11. López de Castro, J. A. *Immun. Today* **10**, 239–246 (1989).
12. Winter, G. & Fields, S. *Virology* **114**, 423–428 (1981).
13. Huet, S. *et al.* *Int. Immunol.* **2**, 311–316 (1990).
14. Melton, D. A. *et al.* *Nucleic Acids Res.* **12**, 7035–7056 (1984).
15. Barnstable, C. J. *et al.* *Cell* **14**, 9–20 (1978).
16. Taylor, P., Davey, J., Howland, K., Rothbard, J. & Askonas, B. *Immunogenetics* **26**, 267–272 (1987).
17. Ljunggren, H.-G. *et al.* *Nature* **346**, 476–480 (1990).
18. Schumacher, T. N. M. *et al.* *Cell* **62**, 563–567 (1990).
19. Townsend, A. R. M. *et al.* *Cell* **62**, 285–295 (1990).
20. Brodsky, F. M., Bodmer, W. F. & Parham, P. *Eur. J. Immun.* **9**, 536–545 (1979).
21. *Promega Protocols and Applications* 43 (Madison, Wisconsin, 1988).
22. Pelham, H. R. B. & Jackson, R. J. *Eur. J. Biochem.* **67**, 247–256 (1976).
23. Jackson, R. J. & Hunt, T. *Meth. Enzym.* **96**, 50–74 (1983).
24. Garoff, H., Simons, K. & Dobberstein, B. *J. molec. Biol.* **124**, 587–600 (1978).
25. Burgert, H.-G. & Kvist, S. *Cell* **41**, 987–997 (1985).
26. Kvist, S., Wiman, K., Claesson, L., Peterson, P. A. & Dobberstein, B. *Cell* **29**, 61–69 (1982).
27. Rothbard, J. B. & Taylor, W. R. *EMBO J.* **7**, 93–100 (1988).

ACKNOWLEDGEMENTS. We thank A. McMichael and J. Rothbard for the gift of the NP 384–394 peptide, E. Weiss and J. Burckhardt for the gifts of cDNAs (HLA-B27 and $\beta 2m$), R. F. Pettersson, U. Eriksson, R. Gabathuler and F. Lévy for critical comments on the manuscript, G. Gyllner and M. Still for technical support, and M. Murelius for help in preparing the manuscript.

Genetic evidence equating *SRY* and the testis-determining factor

Philippe Berta*, J. Ross Hawkins, Andrew H. Sinclair, Anne Taylor, Beatrice L. Griffiths, Peter N. Goodfellow† & Marc Fellous‡

Human Molecular Genetics, Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK

‡ Laboratoire d'Immunogenétique Humaine, INSERM U.276, Institut Pasteur, 25/28 rue du Dr Roux, 75724 Paris Cedex 15, France

THE testis-determining factor gene (*TDF*) lies on the Y chromosome and is responsible for initiating male sex determination. *SRY* is a gene located in the sex-determining region of the human and mouse Y chromosomes and has many of the properties expected for *TDF*^{1–3}. Sex reversal in XY females results from the failure of the testis determination or differentiation pathways. Some XY females, with gonadal dysgenesis, have lost the sex-determining region from the Y chromosome by terminal exchange between the sex chromosomes⁴ or by other deletions⁵. If *SRY* is *TDF*, it would be predicted that some sex-reversed XY females, without Y chromosome deletions, will have suffered mutations in *SRY*. We have tested human XY females and normal XY males for alterations in *SRY* using the single-strand conformation polymorphism assay^{6,7} and subsequent DNA sequencing. A *de novo* mutation was found in the *SRY* gene of one XY female; this mutation was not present in the patient's normal father and brother. A second variant was found in the *SRY* gene of another XY female, but in this case the normal father shared the same alteration. The variant in the second case may be fortuitously associated with, or predisposing towards sex reversal; the *de novo* mutation associated with sex reversal provides compelling evidence that *SRY* is required for male sex determination.

SRY sequences were amplified by the polymerase chain reaction from DNA of XY females, as well as normal male controls. The amplified products were cleaved with restriction enzymes

* Present address: Centre de Recherches de Biochimie Macromoléculaire, CNRS LP8402/INSERM U.249, Route de Mende, 34033 Montpellier Cedex, France.

† To whom correspondence should be addressed.

and run on non-denaturing polyacrylamide gels to detect single-strand conformational polymorphisms (SSCP) (Fig. 1). Three patterns were detected: one pattern common to all of 50 normal male controls and the majority of patient samples; a second pattern for XY female individual (A.A.) and a third pattern for XY female (J.N.). The normal father and brother of A.A. were tested and found to give the same pattern as the controls, implying that A.A. has suffered a *de novo* mutation. In contrast, DNA from father (N.) and his XY daughter (J.N.) showed the same pattern. The amplified products (609 base pairs) from both families were cloned and sequenced (Fig. 2a). As predicted by the SSCP assay, A.A. has a *de novo* mutation G → A that is not shared with her father and brother. For J.N., a G → C change was shared with her father. Other than these observed base changes, the XY females J.N. and A.A. had complete sequence identity to *SRY* over the amplified 609 base pairs (bp) (Fig. 2a). Paternity in both families was confirmed by Southern blotting with minisatellite probes^{8,9}.

The *de novo* mutation in A.A. causes a conservative change from methionine to isoleucine at a residue that lies within the putative DNA-binding motif of *SRY* and is identical in all *SRY* and *SRY*-related genes (Fig. 2b). The association of a *de novo* mutation with a new phenotype provides evidence that the phenotype and mutation are related, and by inference that *SRY* is required for male sex determination. Formally this does not exclude the possibility that other genes on the Y chromosome are required for sex determination, but previous work suggests that, if they exist, these genes must be located adjacent to the pseudoautosomal boundary, close to *SRY*^{1,10}.

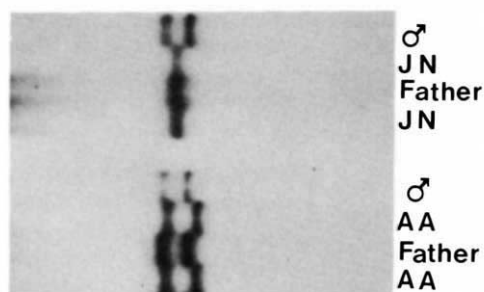


FIG. 1 SSCP analysis of *SRY* in XY females. On the left is family N: ♂, DNA from the normal 46XY male cell line PGF¹²; DNA from the father is flanked on each side by DNA of his XY daughter (J.N.). On the right is family A: again DNA from the father is flanked by that of his XY daughter (A.A.). In total, 11 XY females were tested by SSCP. These individuals: P.F., J.A., A.M., R.B., GM2598, A.S., M., A.A., I.D., J.N. and K.L. all have gonadal dysgenesis, are karyotypically normal and positive both for the Y-pseudoautosomal boundary and ZFY. A.A. and J.N. both have pure gonadal dysgenesis: A.A. has a streak gonad with ovarian stroma and no germ cells; J.N. has a cystic gonad with ovarian stroma and no germ cells. The fathers, A. and N. are both normal, fertile males. Fifty normal males were tested as controls and no variation detected.

METHODS. Polymerase chain reactions were performed with the primers XES7 and XES2 located within the *SRY* open reading frame, amplifying a 609-bp fragment. The primer sequences are: XES7, 5'CCC GAATTCGAC-AATGCAATCATATGCTCTCTGC3'; XES2, 5'CTGTAGCGGTCCCGTCTGCGGTG3'. PCRs were performed with ~100 ng of genomic DNA, 200 μM each dNTP, 0.5 μM each primer, 1.5 mM MgCl₂, 10 mM Tris (pH 8.3), 50 mM KCl, 0.01% (w/v) gelatin, 0.25 U of Taq polymerase and 0.5 μl of [α -³²P]dCTP (3,000 Ci mmol⁻¹, 10 mCi ml⁻¹) in a volume of 10 μl. Reactions were cycled for 1.2 min at 94 °C, 1.2 min at 65 °C and 2 min at 72 °C for 35 cycles. One μl of the product was digested with *Hinf*I and *Taq*I in the presence of 4 mM spermidine hydrochloride in a 10 μl volume. The digested DNA was diluted 1:10 in 0.1% SDS, 10 mM EDTA, followed by a 1:2 dilution in 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol. Samples of 1–3 μl were heated at 80 °C for 5 min to denature the DNA, then loaded onto 6% acrylamide, 10% glycerol non-denaturing gels using a sequencing-gel apparatus. Electrophoresis was carried out at 25 mA, with a fan heater set on cold directed at the gel as a cooling device. Autoradiography of the dried gels was for 3 days without an intensifying screen.

The variant found in J.N. causes a conservative change from valine to leucine (Fig. 2b). This residue is conserved amongst *SRY* and *SRY*-related sequences with the exception of a mouse autosomal gene (autosomal-4), which has a conservative change to isoleucine at this position². There are several possible explanations for this variant. First, the variant could cause conditional sex reversal depending on other genetic or environmental factors. A similar observation has been described in the mouse, where the ability of some alleles of *Tdy* to induce testis formation depends on the genetic background¹¹. Second, the variant could be fortuitously found in a family segregating for an autosomal or X-linked sex reversing gene. Finally, the variant could cause sex reversal and the father is mosaic for wild-type and variant sequences.

The *SRY* gene of the majority of the XY females we have tested appears normal by the SSCP assay. It is possible that these individuals have mutations in *SRY* that are not detected by the assay we have used, either because they do not cause a band shift or because they fall outside the region tested. Alternatively, these individuals may have mutations in another part of the sex-determining pathway.

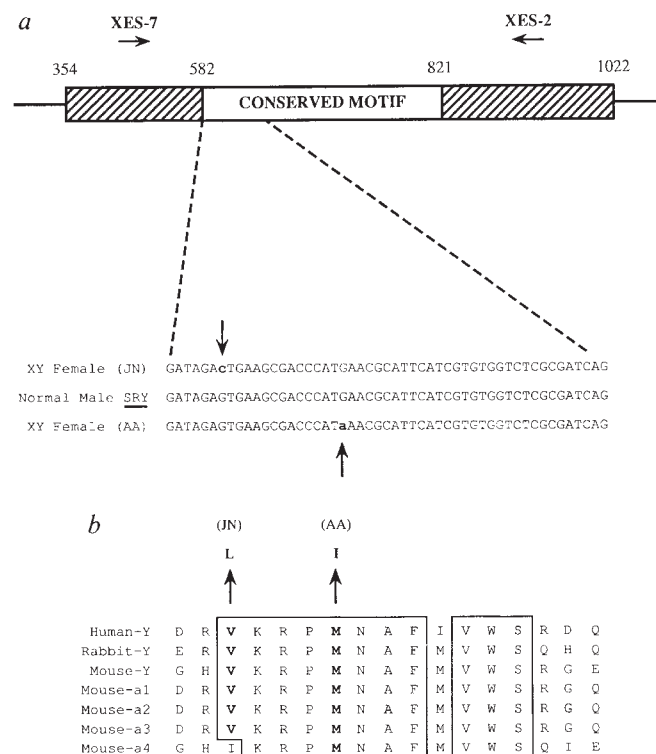


FIG. 2a, The open reading frame of the genomic clone pY53.3 (*SRY*) is shown extending from 354–1,022 bp¹. The conserved motif, which encodes a potential DNA-binding protein, extends from 582 bp to 821 bp¹. The primers XES-7 and XES-2 were used to amplify a 609-bp region. The entire amplification product was sequenced, but only the small region containing the changes in the XY females is displayed. The dotted line indicates the location within the open reading frame of the nucleotide sequence shown. Top line, nucleotide sequence for XY female (J.N.) with base changes from G → C indicated by arrow. Middle line, normal male *SRY* nucleotide sequence. Bottom line, XY female (A.A.) with a base change from G → A indicated by arrow. **b**, The amino acid sequence of *SRY* for the human-Y, rabbit-Y, mouse-Y and mouse autosomal *SRY*-like genes a₁, a₂, a₃, a₄^{1,2}. Boxed shaded regions show identical amino acids conserved across these species. In the XY female J.N. the variant causes a change from valine to leucine, whereas in the XY female A.A. the mutation causes a change from methionine to isoleucine.

METHODS. PCR products were subcloned in pUC18 vectors (NEB). Six independent subclones from individuals (A.A.) and (J.N.) were sequenced on both strands. Double-stranded DNA was sequenced by the dideoxy chain termination method¹³ using synthetic oligonucleotide primers and Sequenase (USB).

In conclusion, a *de novo* mutation in the gene *SRY* is associated with sex reversal in an XY female. This provides compelling evidence that *SRY* is required for testis formation and male sex determination. □

Received 28 September; accepted 18 October 1990.

1. Sinclair, A. H. *et al.* *Nature* **346**, 240–244 (1990).
2. Gubbay, J. *et al.* *Nature* **346**, 245–250 (1990).
3. Koopman, P., Münsterberg, A., Capel, B., Vivian, N. & Lovell-Badge, R. *Nature* **348**, 350–352 (1990).
4. Leveillers, K., Quack, B., Weissenbach, J. & Petit, C. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2296–2300 (1989).
5. Page, D. C., Fisher, E. M. C., McGilivray, B. & Brown, L. G. *Nature* **346**, 279–281 (1990).
6. Orita, M., Suzuki, Y., Sekiya, T. & Hayashi, K. *Genomics* **5**, 874–879 (1989).
7. Orita, M., Iwahana, H., Kanazawa, H., Hayashi, K. & Sekiya, T. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2766–2770 (1989).
8. Wong, Z., Wilson, V., Patel, I., Povey, S. & Jeffreys, A. J. *Ann. hum. Genet.* **51**, 269–288 (1987).
9. Wong, Z., Wilson, V., Jeffreys, A. J. & Thein, S. L. *Nucleic Acids Res.* **14**, 4605–4616 (1986).
10. Palmer, M. S. *et al.* *Nature* **342**, 937–939 (1989).
11. Eicher, E. M. & Washburn, L. L. *A. Rev. Genet.* **20**, 327–360 (1986).
12. Goodfellow, P. J. *et al.* *Ann. hum. Genet.* **53**, 15–22 (1989).
13. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).

ACKNOWLEDGEMENTS. We thank A. Jeffreys for the minisatellite probes, J. P. Raymond, A. Moss, M. Poissonnier for access to the A and N families; N. Abbas and C. Cotinot for technical assistance; J. Weissenbach and R. Lovell-Badge for useful comments and C. Middlemiss for typing the manuscript.

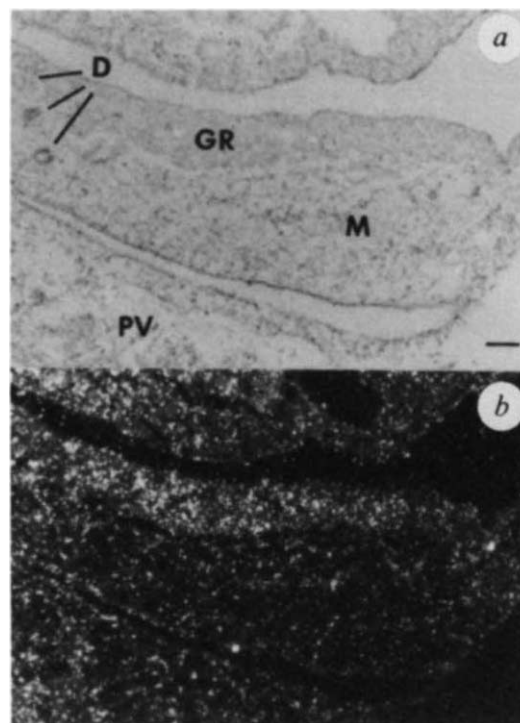


FIG. 2 *In situ* hybridization to a sagittal section of an 11.5-d.p.c. male embryo, using an anti-sense RNA probe for *Sry*. *a*, Bright-field illumination. *b*, Dark-field illumination. GR, genital ridge; M, mesonephros; D, mesonephric and paramesonephric ducts; PV, prevertebrae. Scale bar, 50 μ m. **METHODS.** *In situ* hybridization was carried out as described²². Embryos were sexed by staining for sex chromatin in amnion cells²³. [³⁵S]UTP-labelled anti-sense RNA probe, corresponding to the 374-bp *Bgl*III-*Pst* I fragment of p. 4.2.2. (ref.10) was used. This probe showed no specific labelling to sections of female embryos, and control sense RNA probes did not show a specific signal on male and female sections (data not shown). Autoradiographic exposure was at 4 °C for 6 days.

Expression of a candidate sex-determining gene during mouse testis differentiation

Peter Koopman, Andrea Münsterberg, Blanche Capel, Nigel Vivian & Robin Lovell-Badge

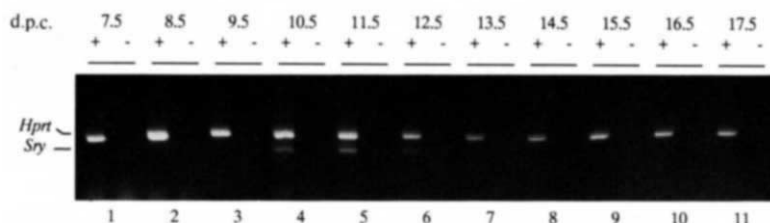
Laboratory of Eukaryotic Molecular Genetics, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

THE development of a eutherian mammal as a male is a consequence of testis formation in the embryo, which is thought to be initiated by a gene on the Y chromosome. In the absence of this gene, ovaries are formed and female characteristics develop¹. Sex determination therefore hinges on the action of this testis-determining gene, known as *Tdy* in mice and *TDF* in humans^{2,3}. In the past, several genes proposed as candidates for *Tdy/TDF* have subsequently been dismissed on the grounds of inappropriate location or expression^{4–9}. We have recently described a candidate for *Tdy*, which maps to the minimum sex-determining region of the mouse Y chromosome^{10,11}. To examine further the involvement of this gene, *Sry*, in testis development, we have studied its expression in detail. Fetal expression of *Sry* is limited to the period in which testes begin to form. This expression is confined to gonadal tissue and does not require the presence of germ cells. Our observations strongly support a primary role for *Sry* in mouse sex determination.

FIG. 1 Time course of fetal *Sry* expression. RNA from mouse embryos of various stages was added to a 'reverse transcription' reaction in the presence (+) or absence (–) of reverse transcriptase (RT). Subsequent PCR reactions contained oligonucleotide primers for hypoxanthine phosphoribosyltransferase (*Hprt*) and *Sry*. As expected, the 352-base-pair (bp) control *Hprt* band was seen in all +RT samples. The 266-bp band corresponding to *Sry* was only seen in 10.5–12.5-d.p.c. samples. As the *Sry* primers are capable of amplifying genomic DNA sequences, the absence of bands from all –RT samples confirms that any signal is due to *Sry* transcripts and not to DNA contamination.

METHODS. Small-scale RNA preparations were made, reverse transcribed and amplified as previously described^{9,10}, with the addition of Perfect Match (Stratagene) to PCR reactions. *Sry* primers 5'–3' were GAGAGC ATGGAG GGCAT and CCACTC CTCTGT GACACT. Annealing was at 53 °C. Samples, divided into + and – RT fractions, were from: 1, twelve whole embryos at

Male and female mouse embryos are morphologically indistinguishable until about 11.5 days post coitum (d.p.c.). The first visible sign of male development, the formation of testes from the 'indifferent' genital ridges, occurs within 24 hours, with the alignment of Sertoli cells into cords. Any candidate for *Tdy* should therefore be expressed in male genital ridge at or before this stage of development. Our initial observations suggested this to be the case for *Sry* (ref. 10). We wished to determine whether *Sry* transcription in the genital ridge is part of a broader temporal or spatial profile of expression suggesting a general role, or whether it correlates more specifically with testis



7.5 d.p.c.; 2, ten whole embryos at 8.5 d.p.c.; 3, 0.1 μ g of poly(A)⁺ RNA from pooled 9.5-d.p.c. embryos posterior to the forelimb-bud; 4, urogenital ridges from six 10.5-d.p.c. embryos; 5, urogenital ridges from a single 11.5-d.p.c. embryo; 6, single 12.5-d.p.c. testis; 7–11, 30, 25, 20, 15 and 12% of single testes from 13.5–17.5-d.p.c. fetuses, respectively. Parkes outbred mice were used unless otherwise stated.

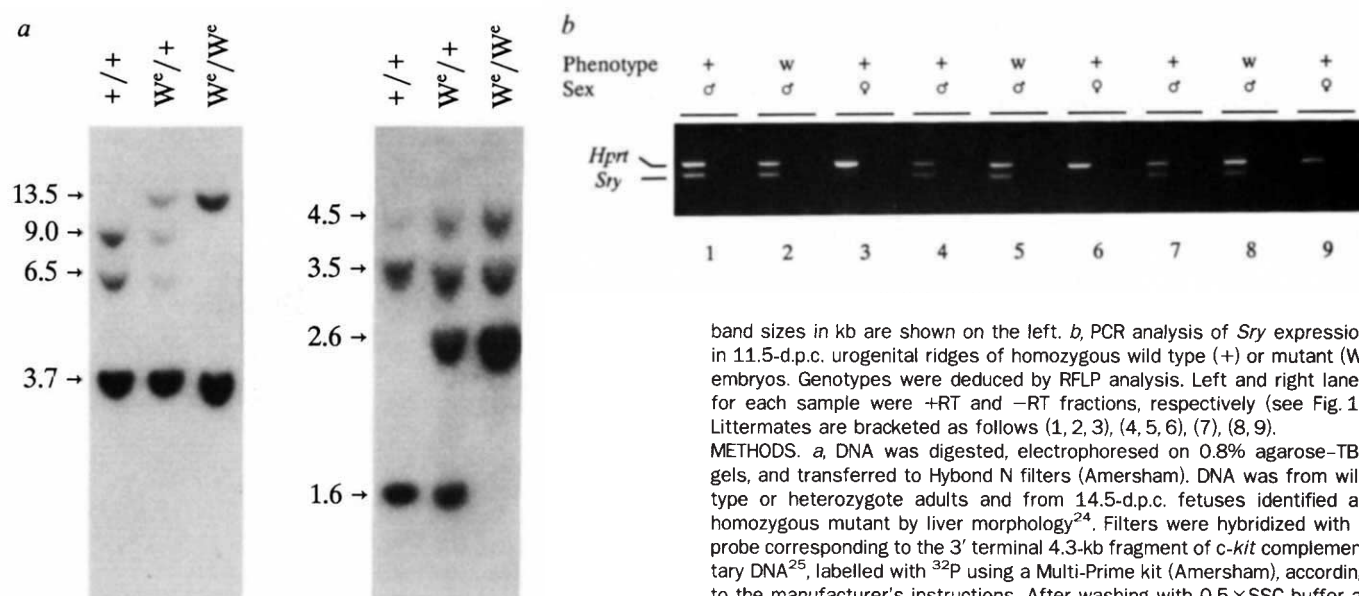


FIG. 3 Analysis of W^e embryos. **a**, Restriction fragment length polymorphism (RFLP) seen on Southern blots of DNA digested with *Bgl*II (left panel) or *Xba*I and *Pst*I (right panel) using a probe for *c-kit* (refs 16, 17); approximate

band sizes in kb are shown on the left. **b**, PCR analysis of *Sry* expression in 11.5-d.p.c. urogenital ridges of homozygous wild type (+) or mutant (W) embryos. Genotypes were deduced by RFLP analysis. Left and right lanes for each sample were +RT and -RT fractions, respectively (see Fig. 1). Littermates are bracketed as follows (1, 2, 3), (4, 5, 6), (7), (8, 9).

METHODS. **a**, DNA was digested, electrophoresed on 0.8% agarose-TBE gels, and transferred to Hybond N filters (Amersham). DNA was from wild type or heterozygote adults and from 14.5-d.p.c. fetuses identified as homozygous mutant by liver morphology²⁴. Filters were hybridized with a probe corresponding to the 3' terminal 4.3-kb fragment of *c-kit* complementary DNA²⁵, labelled with ³²P using a Multi-Prime kit (Amersham), according to the manufacturer's instructions. After washing with 0.5 × SSC buffer at 65 °C, filters were autoradiographed for 16 h. Band sizes were estimated using λ -HindIII markers. **b**, Urogenital ridge RNA from individual progeny of C3H/H-101 W^e /+ matings was divided into +RT and -RT fractions and processed as described in Fig. 1 legend.

differentiation. In particular, we wished to determine whether *Sry* expression in the genital ridge is due to somatic or to germ cells, as it is known that the latter are not required for testis development^{12,13}.

RNA was extracted from mouse embryos and analysed for the presence of *Sry* transcripts by reverse transcription and polymerase chain reaction (PCR). No *Sry* expression was detected in 7.5-, 8.5- and 9.5-d.p.c. embryos, before genital ridge formation. Transcripts were first detected in the genital ridge at 10.5 d.p.c., were present at similar levels in the 11.5-d.p.c. urogenital ridge, and were less abundant in 12.5-d.p.c. testis (Fig. 1). *Sry* transcripts were not detected in testes from 13.5 to

17.5 d.p.c. Therefore, *Sry* expression corresponded precisely with the onset of testis differentiation, with a window of about one day either side.

In situ hybridization was used to examine the distribution of *Sry* transcripts in 11.5-d.p.c. embryos. Although the level of expression was low, the hybridization signal was clearly associated with the genital rather than the adjacent mesonephric component of the urogenital ridge (Fig. 2). The homogeneity of the signal over the genital ridge indicates that the cells expressing *Sry* represent a large population, not confined to any specific region of the genital ridge, and not organized into specific structures. It was not possible to detect hybridization

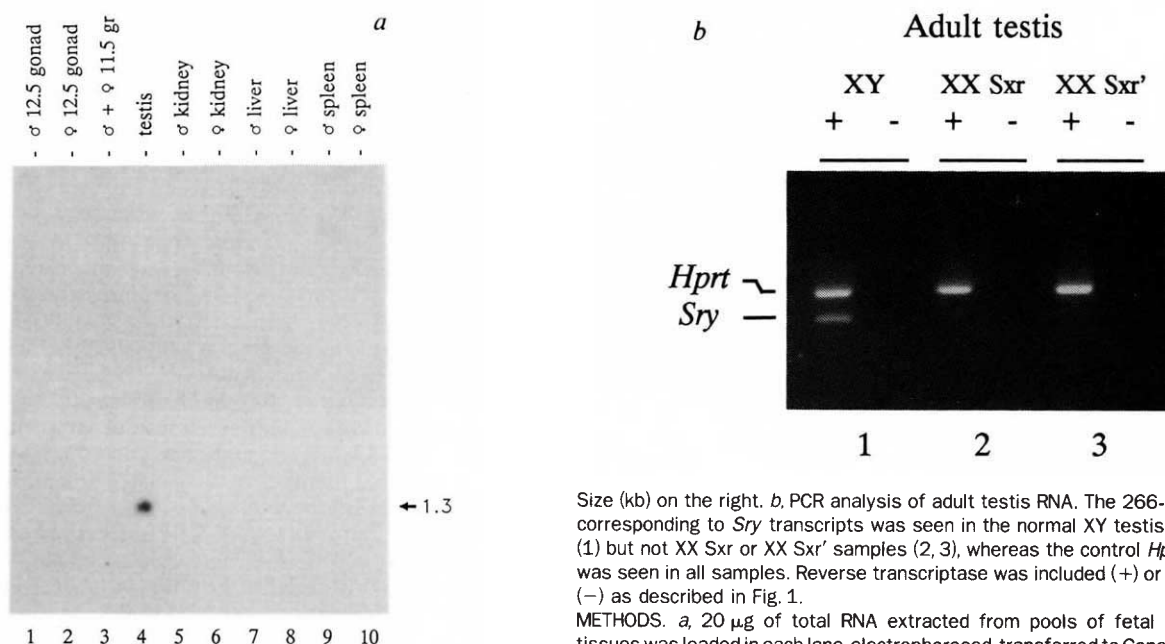


FIG. 4 Analysis of *Sry* expression in fetal and adult tissues. **a**, Northern blot analysis revealed a 1.3-kb transcript in adult XY testis RNA (lane 4), but not in other adult tissues or in 11.5-d.p.c. urogenital ridge (lane 3), 12.5-d.p.c. male (or female) gonads (lanes 1 and 2), or other adult tissues (lanes 5-10).

Size (kb) on the right. **b**, PCR analysis of adult testis RNA. The 266-bp band corresponding to *Sry* transcripts was seen in the normal XY testis sample (1) but not XX Sxr or XX Sxr' samples (2, 3), whereas the control *Hprt* band was seen in all samples. Reverse transcriptase was included (+) or omitted (-) as described in Fig. 1.

METHODS. **a**, 20 μ g of total RNA extracted from pools of fetal or adult tissues was loaded in each lane, electrophoresed, transferred to Gene-Screen (Dupont), and probed with an RNA probe labelled with ³²P to a specific activity of 1×10^9 c.p.m. μ g⁻¹. After hybridization, the filter was treated with RNase to reduce non-specific background, then exposed for 3 days. **b**, Total RNA (0.2 μ g) was divided into +RT and -RT fractions and processed as described in Fig. 1 legend.

above background in 12.5-d.p.c. testes, after the formation of testis cords. The distribution of *Sry* transcripts in fetal tissues at 11.5 d.p.c. was further examined by PCR. When RNA was extracted from urogenital ridge, head, viscera and carcass fractions, there was no evidence for expression in any tissue other than urogenital ridge.

In mouse fetuses homozygous for mutations at the dominant white spotting (*W*) locus, testes form normally despite a lack of germ cells^{12,14}. If *Sry* is to be considered a candidate for *Tdy*, it ought to be expressed in *W/W* as well as in wild-type genital ridges. We were able to identify embryos homozygous for the *W^e* allele¹⁵ by virtue of a restriction fragment length polymorphism detected using a probe for *c-kit*^{16,17} (Fig. 3a). In PCR analyses of *Sry* expression in genital ridge at 11.5 d.p.c., *W^e/W^e* fetuses proved indistinguishable from wild type (Fig. 3b).

We were unable to detect the *Sry* transcripts present in 11.5-d.p.c. genital ridge or 12.5-d.p.c. testis by northern blotting, presumably because of the low level of expression. Northern analysis of adult testis RNA revealed *Sry* transcripts of a similar size to those seen in adult human testis (Fig. 4a). In mouse testis, the transcripts are about 1.3 kilobases (kb) and run as a tight doublet under some conditions. To investigate whether the expression seen in adult testis was somatic, we used PCR to analyse RNA from XX Sxr and XX Sxr' adult mouse testes, which lack germ cells^{18,19}. In contrast to the situation in fetal gonads, *Sry* expression in adult testis is dependent on germ cells (Fig. 4b). We are now addressing the question of whether this expression is a property of germ cells themselves, or a result of somatic-germ cell interaction. It is also unclear whether the adult and fetal transcripts are physically or functionally distinct. A number of genes suspected of having a role in directing developmental events in the embryo show reactivation in adult testis, but in no case has their function there been defined²⁰.

Sry has been shown to meet several criteria that can be applied to any candidate for *Tdy*. It maps to the appropriate area of the Y chromosome, is conserved in a wide range of mammals, is deleted in a line of XY female mice known to be mutant in *Tdy*, and encodes a product consistent with a regulatory role^{10,11,21}. The present results show that *Sry* follows a tightly controlled spatial and temporal profile of expression that correlates well with testis differentiation. As this did not depend on the presence of germ cells, we conclude that *Sry* is expressed in one of the somatic cell lineages present in the developing gonad, as must be the case if *Sry* is to be considered a candidate for *Tdy*. The rapid cessation of *Sry* transcription after testis cord formation suggests that *Sry* initiates a cascade of gene expression, but is not required for the maintenance of gene activity in the developing testis. It will be interesting to determine the nature of the genes involved in subsequent steps of the pathway. □

Received 25 September; accepted 12 October 1990.

- Jost, A., Vigier, B., Prépain, J. & Perchellet, J. P. *Recent Prog. Horm. Res.* **29**, 1-41 (1973).
- McLaren, A. *Trends Genet.* **4**, 153-157 (1988).
- Goodfellow, P. N. & Darling, S. M. *Development* **102**, 251-258 (1988).
- Wachtel, S. S., Ohno, S., Koo, G. C. & Boyse, E. A. *Nature* **257**, 235-236 (1975).
- McLaren, A., Simpson, E., Tomonari, K., Chandler, P. & Hogg, H. *Nature* **312**, 552-555 (1984).
- Singh, L., Phillips, C. & Jones, K. W. *Cell* **36**, 111-120 (1984).
- Page, D. C. et al. *Cell* **51**, 1091-1104 (1987).
- Palmer, M. S. et al. *Nature* **342**, 937-939 (1989).
- Koopman, P. et al. *Nature* **342**, 940-942 (1989).
- Gubbay, J. et al. *Nature* **346**, 245-250 (1990).
- Sinclair, A. H. et al. *Nature* **346**, 240-244 (1990).
- McLaren, A. in *The Origin and Evolution of Sex* (eds Halvorson H. O. & Monroy, A.) 289-300 (Liss, New York, 1985).
- Merchant, H. *Dev. Biol.* **44**, 1-21 (1975).
- Mintz, B. & Russel, E. S. *J. exp. Zool.* **134**, 207-237 (1957).
- Cattanach, B. M. *Mouse News Lett.* **59**, 18 (1978).
- Chabot, B., Stephenson, D. A., Chapman, V. M., Besmer, P. & Bernstein, A. *Nature* **335**, 88-89 (1988).
- Geissler, E. N., Ryan, M. A. & Housman, D. E. *Cell* **55**, 185-192 (1988).
- Cattanach, B. M., Pollard, C. E. & Hawkes, S. G. *Cytogenetics* **10**, 318-337 (1971).
- Burgoyne, P. S., Levy, E. R. & McLaren, A. *Nature* **320**, 170-172 (1986).
- Wolgemuth, D. J. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5813-5817 (1987).
- Lovell-Badge, R. H. & Robertson, E. *Development* **109**, 635-646 (1990).
- Wilkinson, D. G., Bailes, J. A. & McMahon, A. P. *Cell* **50**, 79-88 (1987).

- Monk, M. & McLaren, A. *J. Embryol. exp. Morph.* **63**, 75-84 (1981).
- Borghese, E. *Acta Anat.* **36**, 185-220 (1959).
- Qiu, F. et al. *EMBO J.* **7**, 1003-1011 (1988).

ACKNOWLEDGEMENTS. We thank A. McLaren and M. Buehr for allowing us the use of *W^e* mouse stocks, P. Besmer for the *c-kit* probe, P. Goodfellow for critical comments, D. Wilkinson for assistance with *in situ* hybridization, and J. Gubbay, J. Collignon and A. Economou for stimulating discussions. Also, the photography and oligonucleotide synthesis services at MRC-NIMR for their efficiency. A.M. is supported by the DAAD and Carl Duisberg Stiftung für das Auslandsstudium deutscher Studenten, and B.C. by the NIH.

A human XY female with a frame shift mutation in the candidate testis-determining gene *SRY*

Ralf J. Jäger*, Maria Anvret†, Kerstin Hall‡ & Gerd Scherer*§

* Institute of Human Genetics, University of Freiburg, Breisacherstrasse 33, W-7800 Freiburg i.Br., Germany

† Department of Clinical Genetics and ‡ Department of Endocrinology, Karolinska Hospital, PO Box 60500, S-10401 Stockholm, Sweden

THE primary decision about male or female sexual development of the human embryo depends on the presence of the Y chromosome^{1,2}, more specifically on a gene on the Y chromosome encoding a testis-determining factor, TDF. The human sex-determining region has been delimited to a 35-kilobase interval near the Y pseudoautosomal boundary³. In this region there is a candidate gene for TDF, termed *SRY*, which is conserved and specific to the Y chromosome in all mammals tested³. The corresponding gene from the mouse Y chromosome is deleted in a line of XY female mutant mice, and is expressed at the expected stage during male gonadal development⁴. We have now identified a mutation in *SRY* in one out of 12 sex-inversed XY females with gonadal dysgenesis who do not lack large segments of the short arm of the Y chromosome⁵⁻⁸. The four-nucleotide deletion occurs in a sequence of *SRY* encoding a conserved DNA-binding motif and results in a frame shift presumably leading to a non-functional protein. The mutation occurred *de novo*, because the father of the sporadic XY female that bears it has the normal sequence at the corresponding position. These results provide strong evidence for *SRY* being TDF.

We have characterized a sample of XY females with gonadal dysgenesis for the presence or absence of sequences from interval 1 to 7 of the Y chromosome, including the former TDF candidate gene *ZFY*⁹ (our unpublished data). We selected 10 sporadic and two sib cases to study from this sample, because they did not have deletions in the short arm of the Y chromosome⁵⁻⁸, nor were they familial cases from pedigrees indicating X-linked recessive or male-limited autosomal recessive inheritance¹⁰. These XY females were therefore suited for analysis of mutations in *SRY*.

The sequences of the human *SRY* gene that have been published are from an exon whose limits are not so far precisely known^{3,4}. The open reading frame in this sequence is predicted to encode a sequence with an 80-residue stretch that is characteristic of a DNA-binding motif known as the HMG box (boxed in Fig. 1). This motif, first recognized by comparison of the human transcription activator protein UBF with the high-mobility-group protein HMGI¹¹, is conserved in other or suspected DNA-binding proteins^{3,4,11}. Using oligonucleotide primers flanking the HMG box in the *SRY* sequence, genomic DNA from the 12 XY females and controls was amplified by the polymerase chain reaction (PCR), and the resulting PCR fragments were directly sequenced using two internal sequencing primers as indicated in Fig. 1. In this way, the entire sequence of the 422-base-pair (bp) PCR products could be determined.

§ To whom correspondence should be addressed.

XY female J.S. was diagnosed at age 18 as having primary amenorrhoea owing to hypergonadotropic hypogonadism and absence of secondary sexual characteristics. Her height was 178 cm, and her external genitalia were those of a normal female. Cytogenetic analysis of lymphocytes showed a 46,XY karyotype with no evidence of mosaicism. Gonadal biopsy revealed streak gonads with no signs of malignancy. Histology of excised gonads confirmed the diagnosis of XY gonadal dysgenesis. She has a healthy brother.

A four-base deletion was identified in DNA from J.S. removing nucleotides 773 to 776 near the 3' end of the HMG box motif (Fig. 1). The deletion was present in DNA isolated from cultured cells from a gonadal biopsy (Fig. 2), and in DNA from a skin fibroblast culture of the patient (data not shown). The deletion was in a short dinucleotide repeat AGAGAG. Short repeat sequences are hot spots for mutations^{12,13}. The four-base deletion in J.S. occurred *de novo*, as sequence analysis of DNA from her father showed the normal sequence at the corresponding position (Fig. 2). Paternity was confirmed by a DNA fingerprint of J.S., her father and her mother, using the oligonucleotide (GTG)₅ as a probe¹⁴ (data not shown). The brother of the patient was not available for analysis.

Because of the shift in the reading frame caused by the loss of four nucleotides, the SRY protein of J.S. has a completely different C-terminal sequence from that of normal (Fig. 1). There is no stop codon in the new reading frame up to the end of the published DNA sequences^{3,4}. The sequence of the normal SRY protein is predicted to continue for 68 residues beyond the HMG box motif before reaching a stop codon³. This means that a C-terminal stretch of 84 amino-acid residues of the normal SRY protein is replaced by a new sequence in the SRY protein of

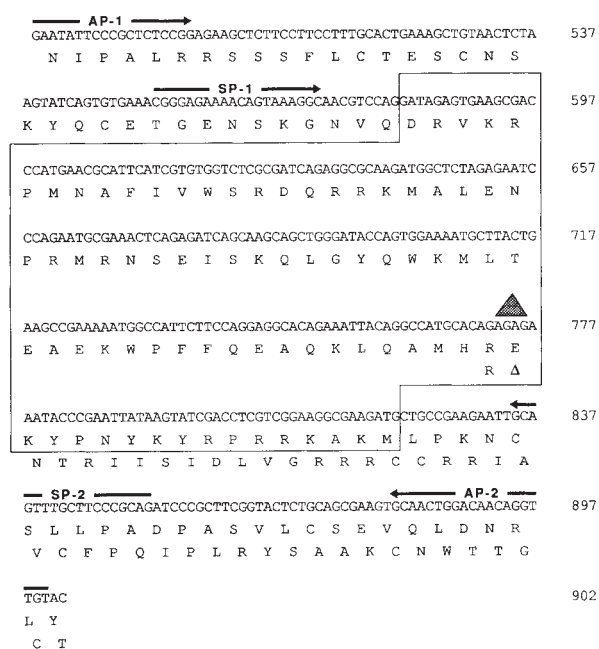


FIG. 1 Nucleotide and corresponding amino-acid sequence of the open reading frame in the human SRY gene. The nucleotide and amino-acid sequence (single-letter code) are from refs 3 and 4, with numbering of nucleotide positions according to ref. 3. The sequence derives from the presumed last exon of the SRY gene. Because the exact limits of this exon have not so far been defined, the 5' stretch of the sequence shown could represent intron sequences. The position of the amplification primers 1 and 2 (AP-1, AP-2) and sequencing primer 1 and 2 (SP-1, SP-2) used for PCR amplification and subsequent sequence analysis are indicated. The HMG box motif comprising 80 amino acids is boxed. The four-nucleotide deletion AGAG in XY female J.S. (Fig. 2) is indicated by the shaded triangle. The amino-acid sequence resulting from the shift in reading frame is given, with the 'Δ' indicating a missing amino-acid residue.

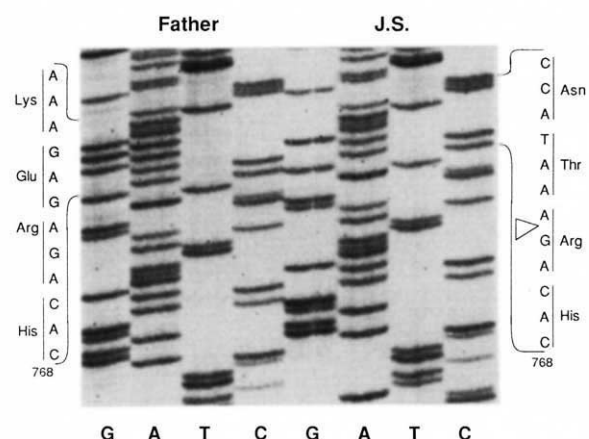


FIG. 2 Four-nucleotide *de novo* deletion in XY female J.S. PCR-amplified DNA was purified and sequenced directly using sequencing primer SP-1 (Fig. 1). The sequence from DNA of the father of XY female J.S. is on the left and from DNA extracted from cultured cells of a gonadal biopsy of XY female J.S. on the right. The nucleotide sequence and the encoded amino acids from the region of interest are shown. The four nucleotides affected by the deletion are shaded on the left and the corresponding position marked by an arrow on the right. G, A, T, C: sequencing lanes.

METHODS: PCR amplification of DNA samples (0.5 µg) was performed in five parallel 50 µl reactions containing 50 mM KCl, 10 mM Tris-HCl, pH 8.4, 3 mM MgCl₂, 0.02 mg ml⁻¹ BSA, 0.05% Tween-20, 0.05% Nonidet P-40, 0.15 mM each dNTP, 0.5 µM each amplification primer, and 1 U Taq polymerase (Amersham). Positions of amplification primers are indicated in Fig. 1, with the sequence AGAATTC and GGAATTC added to the 5' end of primer AP-1 and AP-2, respectively. Reactions were heated for 35 cycles (one cycle: 1 min, 93 °C; 1.5 min, 57 °C; 3 min, 72 °C) in a Techne PHC-1 thermocycler. The five reactions were pooled, and products purified by spin dialysis using a Centricon 100 column (Amicon). Amplified DNA (0.25 pmol) was sequenced directly with sequencing primer SP-1 or SP-2 (10 pmol) (Fig. 1) as described¹⁵, using Sequenase (USB) and ³⁵S-labelled dATP.

J.S. As the coding capacity of the 1.1-kb SRY transcript³ is not likely to exceed 300 amino acids, almost a third of the SRY protein is affected by the frame-shift mutation in J.S. This large change in primary structure, which also affects part of the HMG box motif, the putative DNA-binding domain of the protein, probably leads to a non-functional SRY gene product.

In conclusion, we have identified a *de novo* mutation in the SRY gene associated with sex inversion in a human XY female. This finding complements the observation that a deletion on the mouse Y chromosome that includes the mouse homologue Sry results in sex inversion in mice⁴, and provides additional evidence for SRY being the primary sex-determining gene TDF. But another gene with the same embryonic expression pattern as SRY⁴ could be present in the Y-chromosomal 35-kb region that is critical for human sex-determination³. Until the existence of such an additional gene has been ruled out, mutations in SRY correlating with sex inversion do not strictly prove that SRY and the 'master' testis-determining locus are the same. The other XY females studied with a normal sequence for the HMG box region of SRY could carry inactivating mutations in other parts of the gene, or in a locus flanking SRY. Alternatively, their phenotype results from X-chromosomal or autosomal mutations¹⁰ at loci acting secondary to TDF. □

Received 19 September; accepted 29 October (1990).

1. Ford, C. E. Miller, O. J., Polani, P. E., de Almeida, J. C. & Briggs, J. H. *Lancet* **i**, 711-713 (1959).
2. Jacobs, P. A. & Strong, J. A. *Nature* **183**, 302-303 (1959).
3. Sinclair, A. H. et al. *Nature* **346**, 240-244 (1990).
4. Gubbay, J. et al. *Nature* **346**, 245-250 (1990).
5. Distèche, C. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7841-7844 (1986).
6. Müller, U., Lalonde, M., Donlon, T. Latt, S. A. *Nucleic Acids Res.* **14**, 1325-1340 (1986).
7. Affara, N. A. et al. *Nucleic Acids Res.* **15**, 7325-7342 (1987).
8. Leveilliers, J., Quack, B., Weissenbach, J. & Petit, C. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2296-2300 (1989).

9. Page, D. C. *et al. Cell* **51**, 1091–1104 (1987).
10. Simpson, J. L., Blagowidow, N. & Martin, A. O. *Hum. Genet.* **58**, 91–97 (1981).
11. Jantzen, H.-M., Admon, A., Bell, S. P. & Tjian, R. *Nature* **344**, 830–836 (1990).
12. Farabough, P. J., Schmeissner, U., Hofer, M. & Miller, J. H. *J. molec. Biol.* **126**, 847–857 (1978).
13. Weber, J. L. & May, P. E. *Am. J. hum. Genet.* **44**, 388–396 (1989).
14. Schäfer, R., Zischler, H., Birsner, U., Becker, A. & Epplen, J. T. *Electrophoresis* **9**, 369–374 (1988).
15. Casanova, J.-L., Pannetier, C., Jaulin, C., Jaulin, C. & Kourilsky, P. *Nucleic Acids Res.* **18**, 4028 (1990).

ACKNOWLEDGEMENTS. We (R.J.J. and G.S.) thank Dr J. Lindsten for referring the patient to us, and the patient's parents for collaboration. We also thank Drs M. Fraccaro, H.-J. Gebauer, J. Geraedts, M. Habedank, D. Klingmüller, F. Kohlmann, E. Passarge, R. A. Pfeiffer, J. S. Samra and P. Wieacker for additional XY female patient samples, W. Klein and A. Mayerova for help with DNA fingerprinting and T. F. Wienker and U. Wolf for comments on the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft.

Fitness of RNA virus decreased by Muller's ratchet

Lin Chao

Department of Zoology, University of Maryland, College Park, Maryland 20742, USA

WHY sex exists remains an unsolved problem in biology^{1–3}. If mutations are on the average deleterious, a high mutation rate can account for the evolution of sex⁴. One form of this mutational hypothesis is Muller's ratchet^{5,6}. If the mutation rate is high, mutation-free individuals become rare and they can be lost by genetic drift in small populations. In asexual populations, as Muller⁵ noted, the loss is irreversible and the load of deleterious mutations increases in a ratchet-like manner with the successive loss of the least-mutated individuals. Sex can be advantageous because it increases the fitness of sexual populations by re-creating mutation-free individuals from mutated individuals and stops (or slows) Muller's ratchet. Although Muller's ratchet is an appealing hypothesis, it has been investigated and documented experimentally in only one group of organisms—ciliated protozoa². I initiated a study to examine the role of Muller's ratchet on the evolution of sex in RNA viruses and report here a significant decrease in fitness due to Muller's ratchet in 20 lineages of the RNA bacteriophage $\phi 6$. These results show that deleterious mutations are generated at a sufficiently high rate to advance Muller's ratchet in an RNA virus and that beneficial, backward and compensatory mutations cannot stop the ratchet in the observed range of fitness decrease.

RNA viruses have high spontaneous rates of mutation. Whereas the mutation rate of DNA is on the order of 10^{-9} to 10^{-10} errors per nucleotide replication, the error rate for RNA is 10^{-4} to 10^{-5} (ref. 7). Like most DNA viruses, many RNA viruses can reproduce sexually (exchange genetic material among genomes) when two or more viral particles coinfect the same host cell⁸. In some RNA viruses, the exchange is through the production of hybrid progeny following recombination between the genomes of the coinfecting viruses. In others, the viral genome is segmented into several molecules and hybrid progeny are reassortments of segments from the coinfecting viruses. The bacteriophage $\phi 6$ has a genome that is divided into three RNA segments⁹. Segment exchange may be the primary mechanism for sex in RNA phages because recombination is unknown in these viruses^{10,11}. A nonsegmented RNA phage is effectively asexual.

Twenty lineages were derived from a parent plaque-purified clone of $\phi 6$ and propagated under conditions of intensified genetic drift. The host bacterium *Pseudomonas phaseolicola*¹² served as lawn. Bacterial cultures and phage lysates were prepared with LC broth and agar¹³. Lysates were filtered (0.22 μ m Durapore, Millipore) to remove bacteria and then stored at 4 °C. A lineage was initiated by streaking¹⁴ the parent clone on a plate. After 24 h at 25 °C, a plaque was randomly picked from the plate and streaked onto a new plate. Streaking was used to minimize cross-contamination between lineages, which was a problem with pipettor aerosols. A growth cycle was defined as

TABLE 1 Decrease in fitness of transferred clones

Estimated mean fitness over all 20 transferred clones

$$\begin{aligned}\bar{x} &= 0.78 \\ \Pr [0.671 \leq \bar{x} \leq 0.884] &= 0.95 \\ \text{Range} &= 1.0645 - 0.2853 = 0.7792 \\ t_s &= 4.787 \\ P \text{ (one tail; d.f. = 19)} &= 0.0001\end{aligned}$$

Fitness of transferred clones was measured by paired-growth experiments. Mean ($\bar{x}$) and its 95% confidence limits were calculated¹⁵ from the means of three measurements made on each transferred clone. To approximate better a normal distribution, the mean of each transferred clone was transformed by taking the square root of the absolute value of its logarithm. Values shown here have been retransformed back to fitness. The t -statistic and associated P -value are for a one-tailed hypothesis¹⁵ that the mean fitness over all 20 transferred clones was less than 1, which is the mean fitness of the parent clone. The range shows highest and lowest mean fitness for a transferred clone.

METHODS. Paired-growth experiments. For the first cycle of paired growth, a transferred clone was mixed at a known ratio with a genetically marked parent clone, plated on LC agar with a *P. phaseolicola* lawn, and incubated for 24 h at 25 °C. All plaques on the plate were then collected¹³ and filtered. Two additional cycles of paired growth were made by plating the collected phage and repeating the process. The marker on the parent clone was a host range mutation that allowed growth on the alternate host *P. pseudoalcaligenes*¹². Ratios of transferred and parent clones during paired growth were monitored by plating the collected phage on mixed lawns of *P. phaseolicola* and *P. pseudoalcaligenes* (200:1 ratio), on which the transferred and marked clones made, respectively, turbid and clear plaques. The number of plaques per paired-growth and mixed lawn plate was kept between 200–600 by diluting the collected phage with LC broth. A maximum of 600 was chosen to minimize overlap, and thus genetic exchange, between plaques. W was estimated by fitting log (W) by the least-squares method¹⁵ to log-transformed paired-growth data. A paired-growth experiment mixing the marked and unmarked parent clones showed a small deleterious effect to the marker (average W of unmarked clone = 1.0550; $n = 3$). All fitness values reported were therefore scaled with division by 1.0550. This scaling adjusts for the effects of the marker by setting the mean fitness of the unmarked parent to 1.

the expansion of a lineage within a plaque. During each cycle, the lineage increased from one individual to about 8×10^9 particles per plaque. This protocol intensifies genetic drift by forcing a lineage through a bottleneck of one random individual each growth cycle. After 40 growth cycles, a final random plaque was picked from each of the 20 lineages. Phage isolated from the final plaques were denoted 'transferred' clones.

To determine whether the transferred clones had accumulated deleterious mutations as a result of Muller's ratchet, the fitness of the clones was estimated by paired-growth experiments (Table 1). A transferred clone and the 'parent' clone (see above) were mixed and plated on a *P. phaseolicola* lawn. After incubation for one cycle of paired growth, the resulting phage were collected and passed through two more cycles. The ratio of transferred to parent clone was monitored over the cycles by using a genetically marked parent clone (Table 1). Fitness (W) was defined as the relative change of the transferred:parent clone ratio per paired-growth cycle, or $W^n = R_n/R_0$, where R_0 and R_n are the ratios, respectively, at the start and after n paired-growth cycles. The mean fitness of the parent clone was scaled to 1. A transferred clone with $W < 1$ suffered a loss in fitness. Fitness measurements of all 20 transferred clones were replicated three times.

The fitness of some transferred clones increased by 6%, whereas others decreased by as much as 71% (Table 1). On average, transferred clones incurred a significant drop in fitness; mean fitness over all transferred clones was 0.78 and its 95% confidence limits did not include a value of 1 (Table 1). Analyses of variance¹⁵ indicated a significant heterogeneity in fitness among transferred clones (Table 2). Variation among clones is estimated by the broad sense heritability¹⁶ (h_b^2) of their fitness

TABLE 2 Heterogeneity in fitness of transferred clones

Source	d.f.	Mean square	F	P	Expected mean square*
Among clones	19	0.077	18.91	0.0001	$\sigma_w^2 + k\sigma_a^2$
Within clones	40	0.004			σ_w^2
Total	59				

Fitness measurements of the 20 transferred clones were replicated three times. Analysis of variance¹⁵ was performed on fitness data from all clones and their replicated measurements. Before analysis, fitness values were transformed by taking the square root of the absolute value of their logarithm (see Table 1). The significant heterogeneity indicates genetic variation in fitness among transformed clones.

* $k=3$, the number of times the fitness of each transferred clone was measured. σ_w^2 and σ_a^2 are the parametric variances of fitness within and among transferred clones.

TABLE 3 Estimated variance components of fitness for transferred clones

s_w^2	s_a^2	h_b^2
0.004	0.024	0.86
Pr [$0.732 \leq h_b^2 \leq 0.936$] = 0.95		

Sample variances of fitness among and within clones, s_w^2 and s_a^2 , were estimated¹⁵ from mean squares in Table 2. Broad sense heritability (h_b^2) was computed as $s_a^2/(s_w^2 + s_a^2)$ (ref. 16) and its 95% confidence limits as described by Becker¹⁹.

(Table 3). Because all 20 transferred clones were derived from a single parent clone, $h_b^2=0$ by definition at the start of the experiment. After 40 growth cycles with intensified genetic drift, the fitness of the transferred clones had a significantly higher h_b^2 and 86% of the total variance was among clones (Table 3). Because the transferred clones had a common origin and were randomized with respect to environment during the measurements of fitness, the variance among clones is genetic¹⁶ and necessarily due to mutations that arose after the lineages were created. My results show that Muller's ratchet advanced in the lineages.

Back mutations cannot stop Muller's ratchet because backward mutation rates are anticipated to be less than forward rates¹; although any mutation can change the sequence of a gene (and perhaps harm it), only one back mutation can restore the original sequence. Deleterious mutations may, however, also be offset by mutations that do not change a gene back to its original sequence. If such compensatory mutations are sufficiently common, Muller's ratchet can be stopped¹⁷. The relative frequency of back, compensatory, and beneficial mutations remains an empirical issue. Such offsetting mutations did not stop the ratchet in my experiments. Thus, mutations are on the average deleterious in $\phi 6$ within the observed range of fitness loss.

The fitness decrease caused by Muller's ratchet would have occurred regardless of whether $\phi 6$ is sexual (segmented) or not. The advantage of being sexual accrues only if the transferred clones meet through coinfection. Muller's ratchet can be stopped (or slowed) if reassortment of segments during coinfection recreates a mutation-free individual. If $\phi 6$ were nonsegmented, sex would not be possible (see above) and Muller's ratchet could not be stopped. Stopping Muller's ratchet through segment reassortment, however, requires that the mutations fixed by the transferred clones are on different segments. Whether the latter is true, remains to be determined. The high h_b^2 and wide range of fitness in the transferred clones (Tables 1–3) suggest that the mutations are heterogeneous.

Previous studies show that Muller's ratchet operates in protozoa². Nonetheless, the importance of Muller's ratchet in protozoan evolution is questioned¹⁸ because the small populations

(4–5 individuals) necessary to engage the ratchet may not be realized in nature. For the same reason, Muller's ratchet can be argued to be ineffective in viral populations. Population size without additional information on the history of the population can, however, be misleading. The expansion of the $\phi 6$ lineages to a population size of 8×10^9 did not stop the ratchet when the lineages were also passed through a bottleneck of one individual. If viral transmission in nature is often by a single propagule, Muller's ratchet may be active in populations of wild RNA viruses, despite regular population explosions. It is noteworthy that the correlation between sexuality and dispersal is often presented to support the traditional view that sex evolved for generating variation in heterogeneous environments². My results suggest that the same correlation is predicted by Muller's ratchet. If population bottlenecks are created during dispersal by a few propagules, dispersing lineages have a greater need for sex because they are more prone to Muller's ratchet. □

Received 20 August; accepted 28 September 1990.

- Maynard Smith, J. *The Evolution of Sex* (Cambridge University Press, Cambridge, 1976).
- Bell, G. *Sex and Death in Protozoa. The History of an Obsession* (Cambridge University Press, Cambridge, 1988).
- Michod, R. & Levin, B. R. *The Evolution of Sex. An Examination of Current Ideas* (Sinauer, Sunderland, 1988).
- Kondrashov, A. S. *Nature* **336**, 435–440 (1988).
- Muller, H. J. *Mut. Res.* **1**, 1–9 (1964).
- Felsenstein, J. *Genetics* **78**, 737–756 (1974).
- Holland, J. et al. *Science* **215**, 1577–1585.
- Chao, L. *J. theor. Biol.* **133**, 99–112 (1988).
- Semancik, J. S., Vidaver, A. K. & Van Etten, J. L. *J. molec. Biol.* **78**, 617–625 (1973).
- Horiuchi, K. in *RNA Bacteriophages* (ed. Zinder, N. D.) 29–50 (Cold Spring Harbor Laboratory, New York, 1975).
- Mindich, L., Sinclair, J. F., Levine, D. & Cohen, J. *Virology* **75**, 218–223 (1976).
- Mindich, L., Cohen, J. & Weisburd, M. *J. Bact.* **126**, 177–182 (1976).
- Sinclair, J. F., Cohen, J. & Mindich, L. *Virology* **75**, 198–208 (1976).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning. A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Sokal, R. R. & Rolf, F. J. *Biometry* (Freeman, San Francisco, 1981).
- Falconer, D. S. *Introduction to Quantitative Genetics*, 2nd ed. (Longman, Harlow, 1981).
- Wagner, G. P. & Gabriel, W. *Evolution* **44**, 715–731 (1990).
- Lenski, R. *Science* **248**, 901 (1990).
- Becker, W. A. *Manual of Quantitative Genetics*, 3rd ed. (Washington State University Press, Pullman, 1975).

ACKNOWLEDGEMENTS. I thank C. Matthews and T. Tran for dedicated assistance in measuring viral fitness 7 days a week, D. Carr, R. Fundyga, R. Lenski, B. Levin, D. Gill, and G. Stirling for comments, and L. Mindich and H. Revel for advice, phage and bacterial strains. This work was supported in part by the National Institutes of Health (General Medical Sciences and Biomedical Support Program) and the Office of Graduate Studies and Research at the University of Maryland.

The mitochondrial chaperonin hsp60 is required for its own assembly

Ming Y. Cheng, F.-Ulrich Hartl* & Arthur L. Horwich

Howard Hughes Medical Institute and Department of Human Genetics, Yale School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA

* Institut für Physiologische Chemie der Universität München, Goethestrasse 33, 8000 München 2, Germany

HEATSHOCK protein 60 (hsp60) in the matrix of mitochondria is essential for the folding and assembly of newly imported proteins¹. Hsp60 belongs to a class of structurally related chaperonins found in organelles of endosymbiotic origin and in the bacterial cytosol^{2–9}. Hsp60 monomers form a complex arranged as two stacked 7-mer rings⁶. This 14-mer complex binds unfolded proteins at its surface, then seems to catalyse their folding in an ATP-dependent process¹⁰. The question arises as to how such an assembly machinery is itself folded and assembled. Hsp60 subunits are encoded by a nuclear gene and translated in the cytosol as precursors⁷ which are translocated into mitochondria and proteolytically processed. In both intact cells and isolated mitochondria of the

hsp60-defective yeast mutant *mif4*, self-assembly of newly imported wild-type subunits is not observed. Functional pre-existing hsp60 complex is required in order to form new, assembled, 14-mer. Subunits imported *in vitro* are assembled with a surprisingly fast half-time of 5–10 min, indicative of a catalysed reaction. These findings are further evidence that self-assembly may not be the principal mechanism by which proteins attain their functional conformation in the intact cell.

A high-copy (2μ) yeast plasmid, pGalhsp60, containing the Gal-1 promoter joined with the coding sequence for the wild-type hsp60 precursor, was introduced into the *mif4* strain, which contains a temperature-sensitive lethal mutation in the hsp60 gene¹. At 37 °C (non-permissive temperature) the hsp60 complex in *mif4* loses its functions in folding and assembly of imported mitochondrial proteins. Cell growth ceases within one generation, but the cells resume growth if the temperature is shifted down to 23 °C within 6 hours. We asked whether the growth-deficient phenotype could be rescued by induction of wild-type hsp60 subunits at 37 °C.

Cultures of pGalhsp60/*mif4* and pGalhsp60/wild-type (wt) cells were shifted from 23 °C to 37 °C and two hours later were transferred from ethanol-glycerol medium into galactose-minimal medium to induce the expression of wild-type hsp60. During a 24-hour period after induction, the pGalhsp60/*mif4* culture failed to increase in cell density, in contrast to the pGalhsp60/wt culture (Fig. 1). Strikingly, when expression of wild-type hsp60 was induced by galactose in *mif4* cells for two hours before the shift to 37 °C, the mutant cells grew at ~50% of the rate of wild-type (Fig. 1). Apparently, wild-type complexes assembled at the permissive temperature could rescue the *mif4* phenotype at the nonpermissive temperature.

We determined the fate of the hsp60 subunits in mitochondria isolated from these cells. Immunoblot analysis with anti-hsp60 antiserum was carried out on both soluble and insoluble mitochondrial fractions prepared by centrifugation of Triton extracts. Mutant hsp60 subunits can be distinguished from wild-type subunits by their slower migration in SDS-polyacrylamide gels (Fig. 2; additional data available). As previously observed¹, mutant subunits from noninduced pGalhsp60/*mif4* cells grown

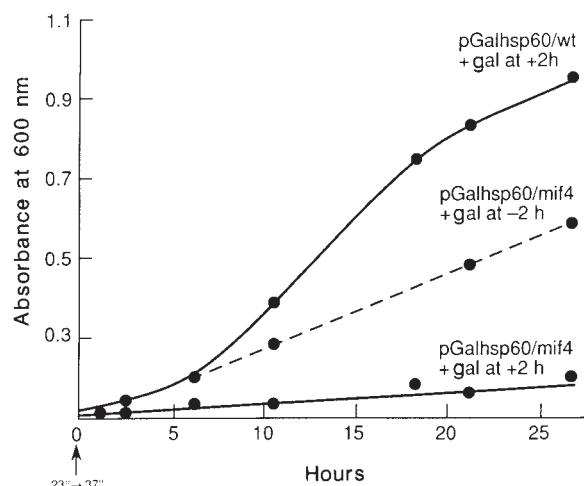


FIG. 1 Growth of wild-type and *mif4* cells after induction of wild-type hsp60 at 37 °C.

METHODS. Galhsp60/wt and Galhsp60/*mif4* strains were grown overnight at 23 °C in minimal medium containing 2% glucose and $20\mu\text{g ml}^{-1}$ of required amino acids (arginine, histidine, leucine), to an absorbance at 600 nm of 0.5. Cells were collected, washed once and resuspended in equal amounts of minimal medium containing 2% ethanol and 3% glycerol (MMEG). Cultures were grown to an absorbance of unity and diluted tenfold in fresh MMEG. After 2 h the temperature was shifted from 23 °C to 37 °C (designated time zero). Galactose was added to a final concentration of 2% either 2 h before (–2 h) or after (+2 h) temperature shift. Growth was monitored by measuring the absorbance of the culture at 600 nm.

at 23 °C were found only in the soluble fraction (lanes 1 and 2), whereas in cells incubated at 37 °C, all the mutant hsp60 subunits were in the pellet fraction (lanes 3 and 4)¹.

When pGalhsp60/*mif4* cells were galactose-induced at 23 °C, soluble wild-type subunits were now evident in the mutant mitochondria (Fig. 2, lanes 5 and 6). But when the mutant strain was induced at 37 °C, both the newly synthesized wild-type subunits and the mutant subunits were found in the insoluble fraction (lanes 7 and 8). Apparently, in *mif4* cells at 37 °C new hsp60 complex cannot be produced, even from wild-type subunits.

If this interpretation is correct, the fate of newly synthesized wild-type hsp60 subunits should depend solely on whether or not functional hsp60 complex is present at the onset of their synthesis. Indeed, when pGalhsp60/*mif4* cells were galactose-induced at 23 °C and then shifted to 37 °C, all of the induced subunits, including those that must have been synthesized after the temperature shift, were found in the soluble fraction (Fig. 2, lanes 9 and 10). Likewise, when a heterozygous diploid strain was grown at 37 °C, the wild-type subunits were in the soluble fraction and the mutant subunits were insoluble (lanes 11 and 12). In both experiments the wild-type subunits fractionated independently of mutant subunits; thus, the respective subunits must have assembled almost exclusively with each other. This independent assembly of wild-type and mutant subunits explains why the *mif4* mutation behaves as a recessive. In each case the assembly of induced wild-type subunits was associated with the presence of hsp60 function at the time of induction. We conclude that pre-existing functionally active hsp60 complex is required to produce new complex from imported subunits.

To investigate this further, we assayed the assembly of wild-type hsp60 subunits imported *in vitro* into mitochondria of wild-type or *mif4* cells. Import of ³⁵S-radiolabelled precursor into mitochondria of both strains and proteolytic processing occurred with comparable efficiency at either 23 °C or 37 °C (Fig. 3a). Mitochondria isolated from the *in vitro* import mixtures were solubilized with Triton X-100 and the soluble fractions applied to a native polyacrylamide gel. Wild-type hsp60 complex migrates to a characteristic position in non-denaturing gels¹⁰. Extracts of wild-type mitochondria which had imported hsp60 precursor at either 23 °C or 37 °C contained a major radiolabelled species that exactly co-electrophoresed with assembled hsp60 complex pre-existent in these mitochondria.

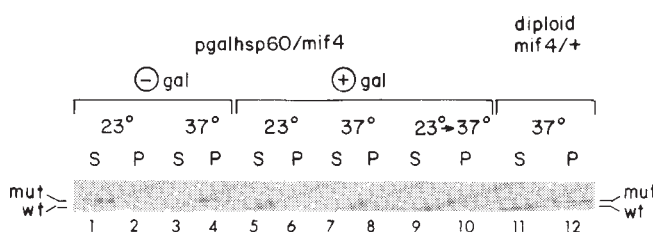


FIG. 2 Immunoblot analysis of hsp60 subunits in mitochondrial extracts. **METHODS.** Yeast strains, including the *mif4* heterozygous diploid¹, were grown at 23 °C in MMEG (see legend to Fig. 1). Five optical density units of cells were collected by centrifugation at 23 °C. pGalhsp60/*mif4* cells were resuspended in 2 ml fresh MMEG and incubated for 4 h at 23 °C or 37 °C, or incubated first for 2 h at 23 °C and then for 2 h at 37 °C. Diploid *mif4* + cells were grown for 4 h at 37 °C. Galactose was added to pGalhsp60/*mif4* cultures to 2% as indicated, either 2 h after commencement of incubation for 23 °C and 37 °C incubations, or at the start of incubation in the temperature-shift experiment. After incubation, cells were collected, spheroplasted in 1 ml Zymolyase solution (1.2 M sorbitol, 20 mM potassium phosphate, pH 7.4, 25 mM 2-mercaptoethanol and 0.2 mg ml⁻¹ Zymolyase 100T), and were lysed in 50 μl of 1% Triton X-100/20 mM HEPES, pH 7.4. After centrifugation (15 min at 12,500 r.p.m. in SS34 rotor), supernatant and pellet fractions were analysed by SDS-PAGE¹³ and immunoblotted with rabbit anti-yeast hsp60 antiserum.

Both newly imported subunits and the pre-existent complex were equally highly resistant to proteinase K (Fig. 3b). This indicates that the imported subunits were not simply associated with the pre-existent hsp60 complex, but had themselves formed an assembled complex. An extract of *mif4* mitochondria which had imported the wild-type hsp60 precursor at 23 °C contained considerably less assembled hsp60 than extracts of wild-type organelles (Fig. 3c). This corresponds with the slow growth of the mutant strain and with its frequent conversion to petite even under these 'permissive' conditions. As expected, when these

mitochondria were incubated after the import reaction for 30 min at 37 °C, the newly assembled radiolabelled wild-type hsp60 remained soluble and fractionated to a characteristic position in a sucrose gradient, while the pre-existent mutant hsp60 complex precipitated (not shown; see Fig. 2, lanes 9–12). This indicates that net assembly of new hsp60 complex had occurred, rather than exchange of subunits with pre-existent complex. When import into *mif4* mitochondria was carried out at 37 °C, no assembled species could be detected (Fig. 3c). Thus, the extent of production of hsp60 complex correlates with the level of hsp60 function present in the target mitochondria.

The *in vitro* import reaction allowed us to estimate the kinetics

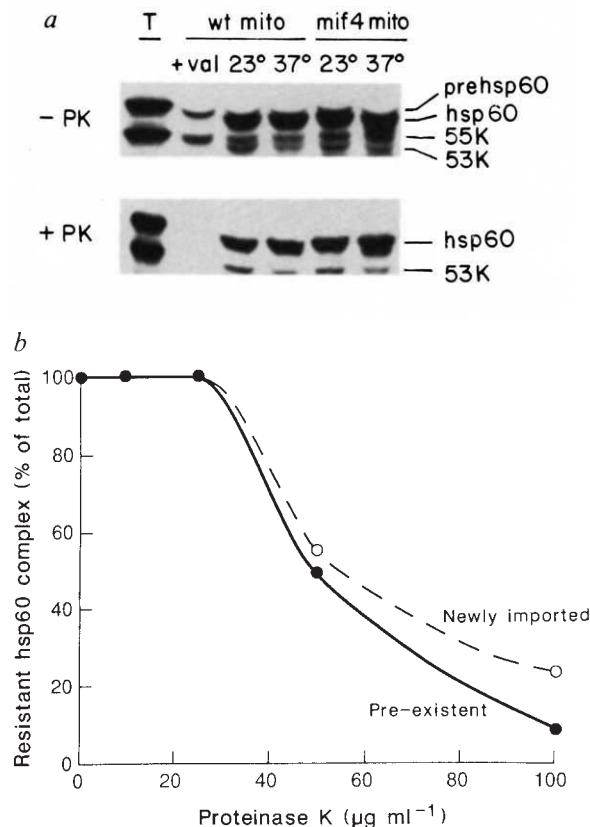


FIG. 3 *a*, Import of hsp60 precursor into isolated mitochondria. *b*, Protease-resistance of imported and pre-existent hsp60. *c*, Assembly of hsp60 imported *in vitro*.

METHODS. *a*, [³⁵S]methionine-labelled precursor of hsp60 was produced by transcription of a plasmid containing a T3 promoter joined with the hsp60 coding region, followed by translation in rabbit reticulocyte lysate^{14,15}. The final translation mix contained 2–3 pmol hsp60 precursor per ml. Import was performed in 10% reticulocyte lysate in translocation buffer (3% bovine serum albumin, 80 mM KCl, 0.25 M sucrose, 1 mM MgCl₂, 10 mM MOPS, pH 7.2) containing 250 μg ml⁻¹ mitochondrial protein¹. NADH (1 mM) and

ATP (1 mM) were included as energy source. In one reaction, 1 μM valinomycin was present during import (+val). After incubation for 30 min at 25 °C or 37 °C, import was stopped by adding valinomycin and the reactions were divided into 3 portions: portions 2 and 3 were treated with proteinase K (PK, 20 μg ml⁻¹) for 20 min at 0 °C. After addition of phenylmethylsulphonyl fluoride to 1 mM, mitochondria were pelleted and portions 1 and 2 analysed by SDS-PAGE (±PK). T, translation mix used in the import reactions (not protease-treated); 55K and 53K designate a secondary translation product and its processed form, respectively. *b*, Wild-type mitochondria treated with proteinase K after import were solubilized at 1 mg ml⁻¹ in 1% Triton X-100, 20 mM HEPES, pH 7.4. After centrifugation at 15,000g for 15 min, the supernatant was divided into 50-μl aliquots which were treated with 0–100 μg ml⁻¹ proteinase K as in *a*. Reactions were halved and then analysed by SDS-PAGE followed either by fluorography or immunoblotting with hsp60 antiserum, respectively¹⁶. Protease-resistant hsp60 was quantified by laser densitometry and is expressed as per cent of total imported or pre-existent hsp60. *c*, Mitochondria from portion 3 of the import reaction in *a* were extracted with Triton (see above). Extracts were treated with 30 μg ml⁻¹ proteinase and then run on native polyacrylamide gels^{9,17} for analysis by fluorography and densitometry. Newly assembled hsp60 co-electrophoresed with pre-existent complex identified by immunoblotting of native gels⁹. The recovery of both newly imported and pre-existent hsp60 on native gels was ~60% as judged by SDS-PAGE of the same reactions. The amount of newly assembled hsp60 detected in wild-type mitochondria is set to 100%.

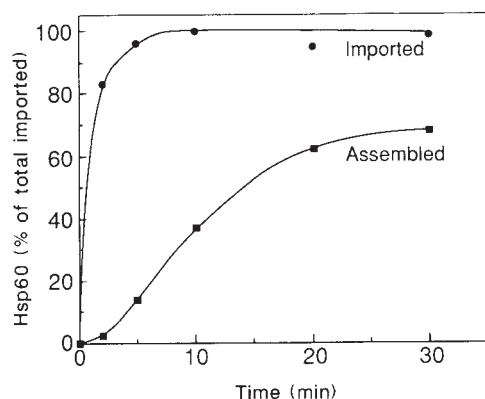


FIG. 4 Kinetics of assembly of newly imported hsp60 subunits.

METHODS. Wild-type mitochondria were incubated for import of radiolabelled hsp60 precursor for up to 30 min (see legend to Fig. 3). After the times indicated, import was terminated by addition of valinomycin and cooling on ice. Mitochondria were treated with protease and the reaction at each time point was divided into two. One was analysed by SDS-PAGE and fluorography (imported hsp60). The other was Triton-extracted and after treatment with 30 μg ml⁻¹ proteinase K was analysed by native gel electrophoresis together with purified hsp60 standard (assembled hsp60). Native gels were stained with Coomassie blue and fluorographed. Amounts of hsp60 are expressed as a per cent of total imported hsp60. One mg mitochondria imported ~2 pmol hsp60 precursor. The protein content of a single mitochondrion is assumed to be 0.1 pg¹⁸.

of assembly of hsp60 subunits into the 14-mer complex. Import was complete after 2 min (Fig. 4), and newly assembled, protease-resistant hsp60 was formed with a half-time of 5–10 min, as judged by native gel electrophoresis. This is surprisingly fast, considering the complex structure of the assembled hsp60 and the low concentration of imported subunits—fewer than 100 copies per mitochondrion. We propose that the ATP-dependent interaction of folding monomers with the hsp60 complex stabilizes a folding intermediate that is rate-limiting for assembly and thereby increases its effective concentration. This would amount to catalysis of assembly.

The biogenesis of the hsp60 complex reflects more generally on the biogenesis and origin of mitochondria. The organelles are not made *de novo* but are produced by growth and division of pre-existent structures¹¹. Our findings extend this principle to the level of the components themselves. The machinery necessary for the folding and assembly of many if not most mitochondrial proteins cannot itself undergo self-assembly. This presumably also applies to the formation of functional groEL complex in the prokaryotic cytosol, and to the chloroplast chaperonin^{3,5,8}.

Mitochondria originated by an endosymbiotic event from a prokaryotic ancestor¹², which must have already contained a groEL complex similar to the one found in present-day bacteria. We can only speculate as to how this prokaryotic complex was itself first assembled. Self-assembly is a possibility, but it seems more likely that the ancestor of groEL formed with the help of

some other factor already in existence, presumably a protein or even a nucleic acid.

Note added in proof: Lissin *et al.*¹⁹ report on the *in vitro* assembly of groEL 14-mer from folded monomers. They conclude that 14-mers are first formed spontaneously and then catalyse further assembly in a Mg-ATP-dependent reaction. In agreement with our results, addition of functional groEL and groES is strongly stimulatory. However, we do not observe 'self-chaperoning' of hsp60 assembly *in vivo* as described for groEL *in vitro*. □

Received 10 September; accepted 19 October 1990.

- Cheng, M. Y. *et al.* *Nature* **337**, 620–625 (1989).
- Laskey, R. A., Honda, B. M., Mills, A. D. & Finch, J. T. *Nature* **275**, 416–420 (1978).
- Ellis, R. J. & Hemmingsen, S. M. *Trends biochem. Sci.* **14**, 339–342 (1989).
- Rothman, J. E. *Cell* **59**, 591–601 (1989).
- Barracough, R. & Ellis, R. J. *Biochim. biophys. Acta* **608**, 19–31 (1980).
- McMullen, T. W. & Hallberg, R. L. *Molec. cell. Biol.* **8**, 371–380 (1988).
- Reading, D. S., Hallberg, R. L. & Myers, A. M. *Nature* **337**, 655–659 (1989).
- Hemmingsen, S. M. *et al.* *Nature* **333**, 330–334 (1988).
- Golubinnoff, P., Christeller, J. T., Gatenby, A. A. & Lorimer, G. H. *Nature* **342**, 884–889 (1989).
- Ostermann, J., Horwich, A. L., Neupert, W. & Hartl, F.-U. *Nature* **341**, 125–130 (1989).
- Luck, D. J. L. *J. Cell Biol.* **24**, 461–470 (1965).
- Hartl, F.-U. & Neupert, W. *Science* **247**, 930–938 (1990).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).
- Cheng, M. Y. thesis, Yale Univ. (1990).
- Pelham, H. R. B. & Jackson, R. J. *Eur. J. Biochem.* **67**, 247–256 (1976).
- Towbin, H., Staehelin, T. & Gordon, J. *Proc. natn. Acad. Sci. U.S.A.* **76**, 4350–4354 (1979).
- Musgrove, J. E., Johnson, R. A. & Ellis, R. J. *Eur. J. Biochem.* **163**, 529–543 (1987).
- Bahr, G. F. & Zeitler, E. *J. Cell Biol.* **15**, 489–501 (1982).
- Lissin, N. M., Vennyaminov, S. Y. & Girschovich, A. S. *Nature* **348**, 339–342 (1990).

ACKNOWLEDGEMENTS. We thank W. Fenton and W. Neupert for discussion. This work was supported by grants from the NIH and Deutsche Forschungsgemeinschaft, by the Böhmalk Research Fund, and by the Bernard and Jennie M. Nelson Fund.

Solution structure of the DNA-binding domain of the oestrogen receptor

John W. R. Schwabe, David Neuhaus & Daniela Rhodes

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

STEROID hormone receptors control gene expression through binding, as dimers, to short palindromic response elements located upstream of the genes they regulate^{1–3}. An independent domain of ~70 amino acids directs this sequence-specific DNA binding and is highly conserved between different receptor proteins and related transcription factors^{4–6}. This domain contains two zinc-binding Cys₂-Cys₂ sequence motifs, which loosely resemble the 'zinc-finger' motifs of TFIIIA^{7–10}. Here we describe the structure of the DNA-binding domain from the oestrogen receptor, as determined by two-dimensional ¹H NMR techniques. The two 'zinc-finger'-like motifs fold to form a single structural domain and are thus distinct from the independently folded units of the TFIIIA-type zinc fingers^{11–13}. The structure consists of two helices perpendicular to each other. A zinc ion, coordinated by four conserved cysteines, holds the base of a loop at the N terminus of each helix. This novel structural domain seems to be a general structure for protein–DNA recognition.

An 84-amino-acid peptide including the DNA-binding domain and the putative nuclear localization domain¹⁴ of the oestrogen receptor (ERDBD, Fig. 1a), was expressed in *Escherichia coli* using the T7 expression system¹⁵. The soluble fraction was purified to near homogeneity. Native DNA-binding activity was demonstrated using quantitative bandshift and DNase I footprinting assays (Fig. 1b). The peptide was shown to be monomeric in solution and each monomer contains two zinc atoms (J.W.R.S. *et al.*, manuscript in preparation).

The purified protein has a limited solubility and stability range and is therefore not an ideal sample for NMR spectroscopy. It was necessary to collect data at pH 6.50 (higher than ideal) and at a protein concentration of about 1 mM (lower than ideal). Protein labelled with ¹⁵N was not prepared because, although

the host cells are prototrophic, cells transformed with the expression plasmid are not viable in minimal medium.

NMR data were collected and assigned according to established procedures^{16,17} (Fig. 2). Despite the limitations described above, these data allowed a complete sequential assignment, with the exception of the three N-terminal residues and the two C-terminal glycines (Fig. 2a), although a few spin-systems remain incomplete, and no signals were assigned for the proline. Details of the sequential assignment will be published elsewhere. (J.W.R.S. *et al.*, manuscript in preparation.)

Assignment of cross-peaks between different residues yields distance constraints which characterize the secondary and tertiary structure of the peptide. A total of 297 inter-residue distance constraints were obtained; 158 are between non-adjacent residues, and 106 of these span four or more residues (Fig. 2b). This relatively low number of distance constraints reflects the less than ideal properties of the ERDBD peptide, but proved sufficient to determine the global fold of the protein.

There has been doubt as to which of the five cysteines coordinate the metal in the second zinc-binding site^{18,19}. Several lines of evidence indicate that Cys 67 is not a zinc ligand. First, short- and medium-range constraints identify a helix from Gln 60 to Met 72 (Fig. 2a). Cys 67 is located on the opposite face of this helix from both Cys 59 and Cys 62, and consequently cannot serve as a zinc ligand. Second, mutagenesis shows that Cys 67 is dispensable for receptor activity, whereas the other cysteines are required¹⁸. Finally, the pattern of long-range nuclear-Overhauser-effect cross-peaks indicates that Cys 67 is involved in structural interactions distant from the metal-binding site. Indeed, preliminary calculations were consistent with there being two independent clusters of zinc ligands, justifying subsequent use of inter-sulphur distance constraints to generate tetrahedral zinc geometry (without imposing a chirality on the arrangement of the sulphur ligands). In addition, stereospecific assignments were made for six out of eight of these cysteine zinc ligands, thereby yielding approximate χ_1 angles for these residues.

Calculations to define the tertiary structure were made using YASAP²⁰, a simulated annealing protocol, within the molecular dynamics program XPLOR²¹. Twenty-eight structures were calculated from random start conformations. Figure 3a shows

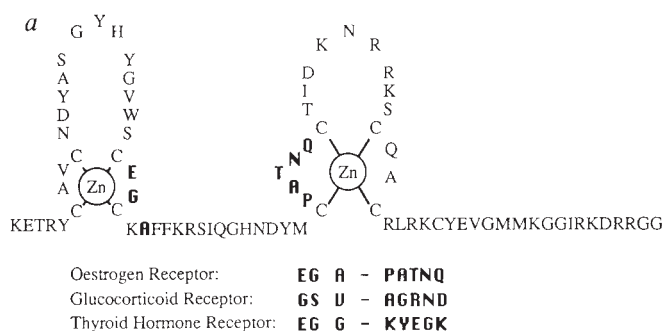


FIG. 1 a, The classical zinc-finger representation of the 84-residue oestrogen receptor DNA-binding domain (the ERDBD) used in this study. Single-letter amino-acid code is used. This region is fully conserved between the chicken and human receptors. Of the five conserved cysteines in the second zinc-binding site, the first four are used to ligate the zinc ion. The residues thought to be involved in discriminating between DNA response elements are shown in bold type, with the corresponding residues in the glucocorticoid and thyroid hormone receptors shown below. **b**, DNase I footprint of the ERDBD-DNA complex. Lane 0, no digestion; Lane G, positions of guanines in the sequence; Lanes 1 and 2, 10- and 20-s digestions of naked DNA; Lanes 3 and 4, 10- and 20-s digestions of ERDBD-DNA complex.

METHODS. A 102-base-pair (bp) fragment (*HindIII/SacI*) containing a wild-type oestrogen response element and flanking sequence (35 bp) was labelled at the *HindIII* site with [α - 32 P]dATP and reverse transcriptase. The DNA (10^{-8} M) was incubated with pure ERDBD (4×10^{-8} M) for 5 min at room temperature. The binding buffer contained 20 mM MES, pH 6.0, 2 mM MgCl₂, 25 μ M zinc acetate, 0.5 M DTT, 3% glycerol and 0.05% Nonidet P-40. DNase I was added to a final concentration of 1 μ g ml⁻¹ for 0, 10 and 20 s. The resulting fragments were analysed on an 8% denaturing polyacrylamide gel.

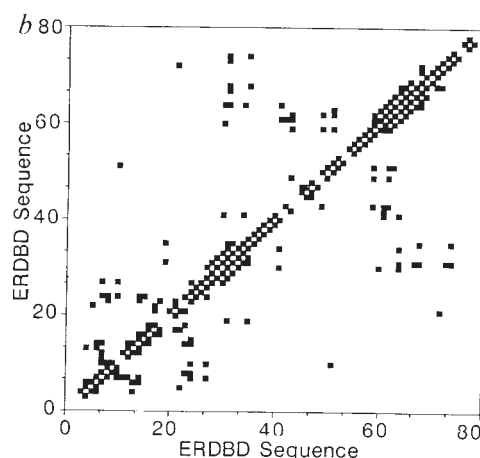
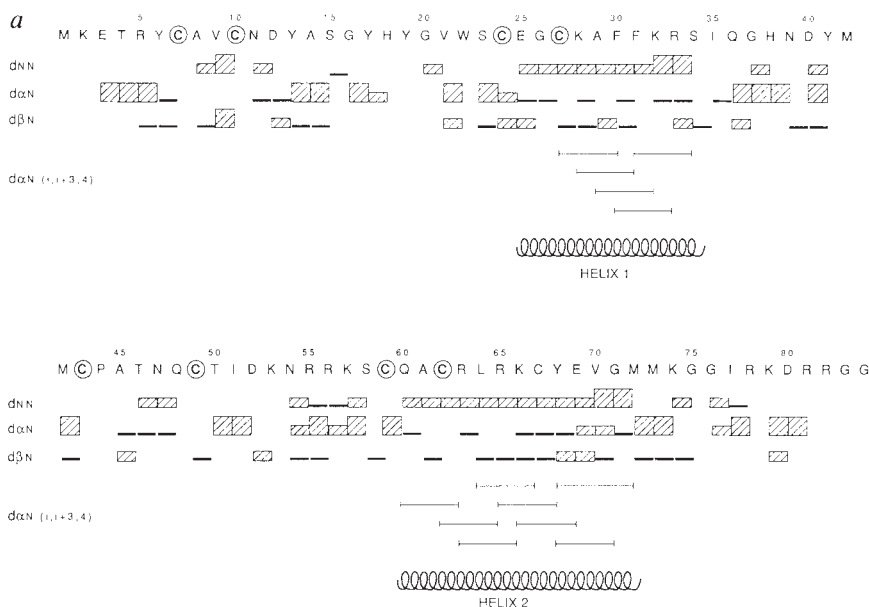
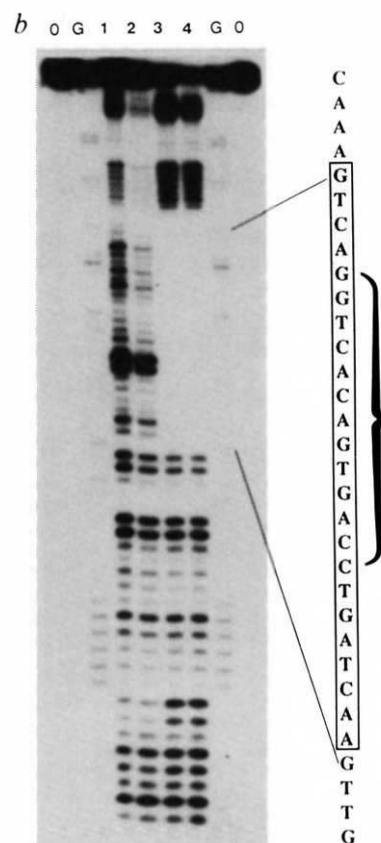


FIG. 2 a, Summary of sequential and medium-range (separated by less than four residues) close contacts in the ERDBD. These data allow identification of secondary structure elements. The location of two helices is shown at the bottom. The size of the blocks roughly represents cross-peak intensity. Cross-peaks seen at mixing times greater than 120 ms are not shown. **b**, Diagonal plot of sequential, medium and long range close contacts in the ERDBD. This plot summarizes the interactions that determine the global fold of the peptide. Each box indicates one or more distance constraints between protons in the indicated amino acids.

METHODS. NMR data were collected using a Bruker 500-MHz AM spectrometer. Samples were prepared in 40 mM pyrophosphate buffer, pH 6.5 at about 1 mM in both 85% H₂O-15% D₂O and 100% D₂O. Experiment times ranged from 18 to 30 h, with 2,048 time domain kpoints $\times$ 512

increments. 2QF-COSY, RELAY, 2Q and TOCSY (τ_m = 40-102 ms, DIPSI-2 mixing sequence²⁶ with z-filters) experiments were used for spin system assignment. Sequential assignments were made using NOESY experiments (τ_m = 200 ms). To resolve ambiguities, TOCSY and NOESY data were collected at four different temperatures (12, 17, 27 and 32 °C). Cross-peak build-up rates were determined using a NOESY τ_m series (30-120 ms) at 17 °C. Intensities were calibrated against intra-residue $d\alpha N$ ($i, i+1$) in extended chain (2.2 Å), dNN ($i, i+1$) in α helix (2.8 Å) and $d\alpha N$ ($i, i+3$) in α helix (3.4 Å). Upper distance bounds were set to 2.5, 3.0 and 3.5 Å. For those peaks that could not be quantified, distance constraints were set at 5 Å. Lower distance bounds were set at the van der Waals' radius. Appropriate pseudo-atom corrections were added to the distance constraints.

a superposition (using a least-squares fit) for 15 of these structures selected on the basis of low residual constraint violations. Residues beyond Gly 75 are not shown owing to the large structural disparity reflecting the absence of non-sequential distance constraints. This region contains the putative nuclear localization domain.

Generally, the structure is well defined except for a loop of nine residues between the cysteine pairs in the second metal-binding site. For seven of these residues there are no long-range distance constraints; indeed NH signals for several of these residues are not present in NOESY spectra above 17 °C, implying that this region may be truly flexible with rapid NH exchange.

The backbone r.m.s. deviation for the superposition of the helical regions (residues 24–35, 59–72; Fig. 3a) is 1.25 Å. The r.m.s. deviation including the N-terminal region (residues 6–35, 59–72) increases to 1.49 Å and to 2.78 Å if the flexible loop is also included (residues 6–72).

A representative structure is shown in Fig. 3b and c. The central feature consists of two amphipathic helices of 11–13 residues packed at ~90° to each other and crossing near their mid-points. The two zinc-binding sites lie at the N termini of each of the two helices. An extensive hydrophobic core is formed between these helices and residues just beyond each of the two helices (Fig. 3b). This core involves Tyr 19*, Val 21, Phe 30*,

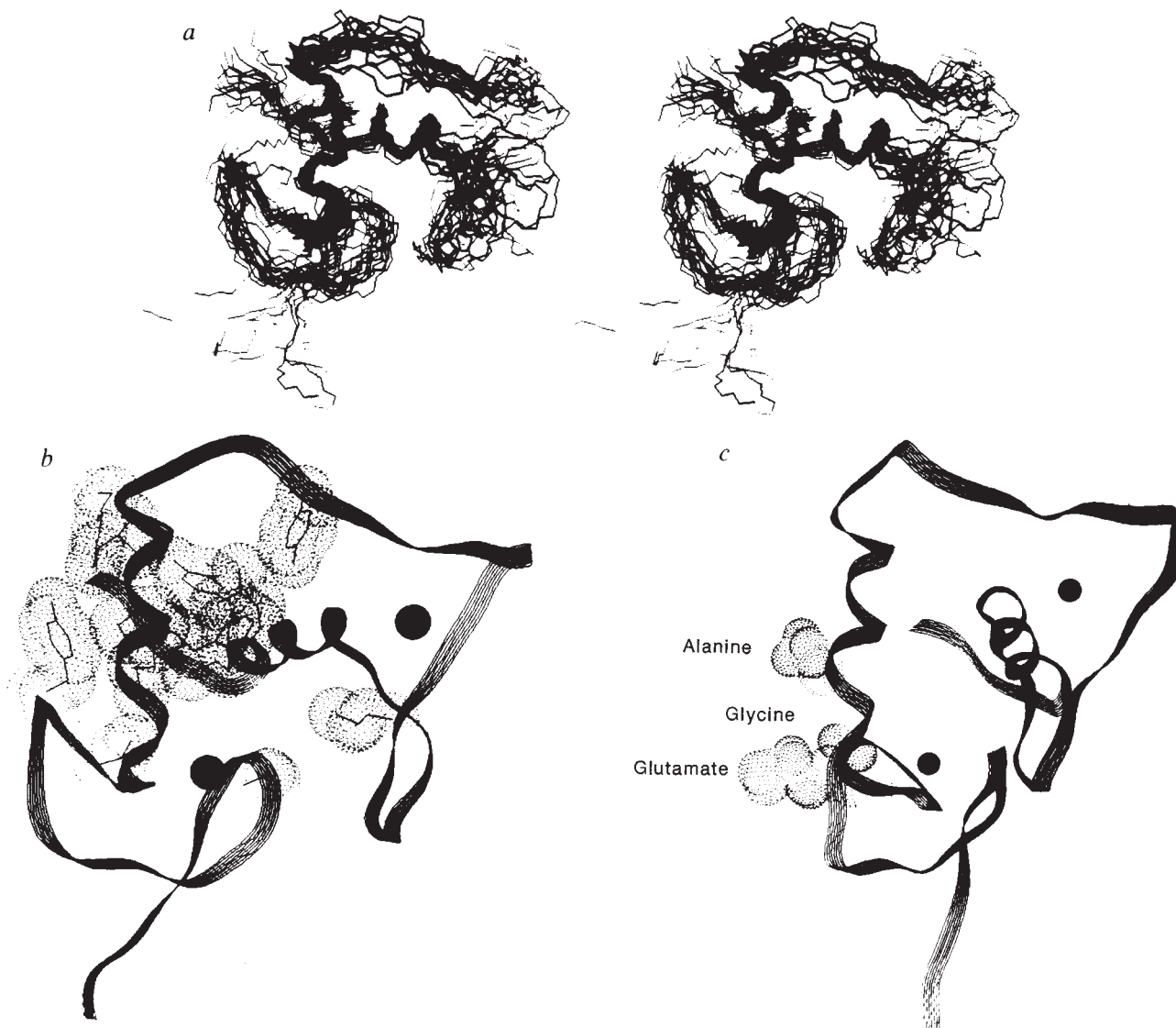


FIG. 3 *a*, Stereo view of an overlay of 15 independently calculated structures selected on the basis of low residual constraint violations, with good geometries. In these selected structures, there are between 1 and 6 violations over 0.4 Å and between 0 and 3 over 0.5 Å. Backbone (C α , C, N) structures are shown superimposed using a least-squares-fit of the α -helical regions (residues 24–35, 59–72) to one of the structures. These structures were calculated using a set of 297 inter-residue distance constraints. No explicit hydrogen-bond constraints were included. Calculations were made using YASAP²⁰, a simulated annealing protocol, within XPLOR, run on a Silicon Graphics Iris 4D/220GTX (about 150 min per structure). There are six stages to the YASAP protocol: (1) generation of a random structure; (2) brief energy minimization; (3) 22.5 ps of dynamics at 1,000 K with nuclear-Overhauser-effect soft-square function and reduced van der Waals' repel function; (4) dynamics with increasing van der Waals' repel function and asymptote for the nuclear-Overhauser-effect soft-square function; (5) cooling to 300 K with square-well nuclear-Overhauser-effect function and further

increasing van der Waals' repel function; (6) final stage of energy minimization. Attractive van der Waals', electrostatic and hydrogen-bonding energy terms were not used during the structure calculations. Zinc atoms and zinc sulphur bonds were subsequently incorporated to emphasize the zinc-binding sites. No further refinement was performed. *b*, Interhelix hydrophobic core. This view is roughly the same as that in Fig. 3a. At least 11 side chains (Tyr 19, Val 21, Phe 30, Phe 31, Ile 35, Tyr 41, Leu 64, Cys 67, Tyr 68, Met 72 and Lys 74) contribute to a hydrophobic core between the two helices. These are shown with van der Waals' dot surfaces (protons not included). Residues Cys 10 and Ile 51 are also highlighted with a dot surface. Zinc atoms are highlighted at 75% van der Waals' radius. *c*, Putative DNA recognition helix. Relative to Fig. 3a, this view is rotated by ~90° clockwise about the first helix. The three amino acids that determine differential specificity (putative DNA-binding residues; Glu 25, Gly 26 and Ala 29) are labelled and shown with van der Waals' dot surface (including protons). Zinc atoms are highlighted at 50% van der Waals' radius.

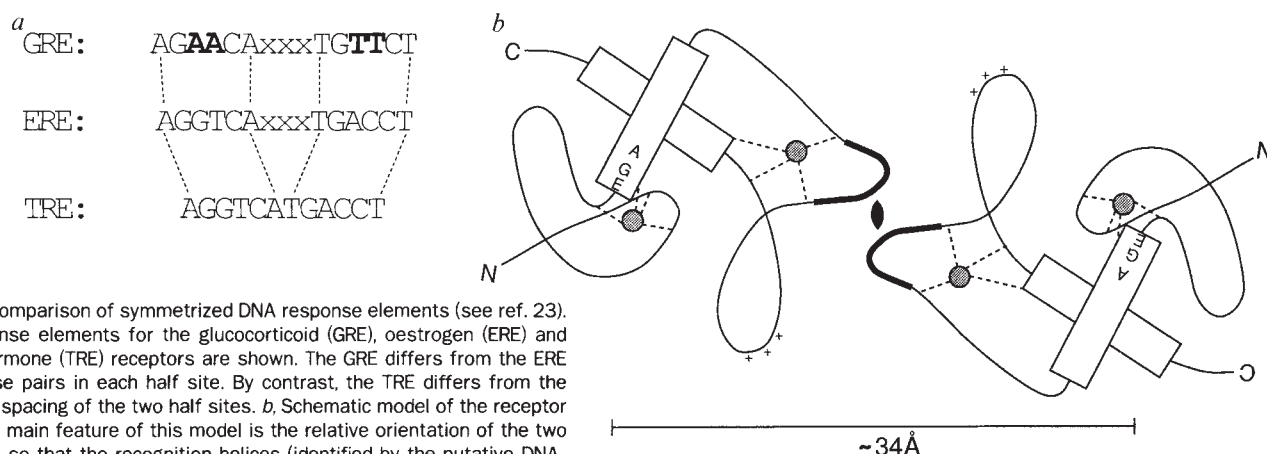


FIG. 4 *a*, Comparison of symmetrized DNA response elements (see ref. 23). The response elements for the glucocorticoid (GRE), oestrogen (ERE) and thyroid hormone (TRE) receptors are shown. The GRE differs from the ERE in two base pairs in each half site. By contrast, the TRE differs from the ERE in the spacing of the two half sites. *b*, Schematic model of the receptor dimer. The main feature of this model is the relative orientation of the two monomers so that the recognition helices (identified by the putative DNA-binding residues Glu (E), Gly (G), Ala (A)) are antiparallel and at a suitable angle to the DNA axis to bind in successive major grooves, 34 Å apart. In addition, conserved basic residues (marked + + +) in the nine-residue loop, at the second zinc-binding site, are in a suitable position to interact with the phosphate backbone of the minor groove. A thickened line indicates the

location of those residues thought to be involved in protein-protein contacts. The distance between the recognition helix and the edge of the structure is about 19 Å as estimated from the NMR structure. The twofold axis between the monomers is marked.

Phe 31*, Ile 35, Tyr 41*, Leu 64, Cys 67*, Tyr 68, Met 72* and Lys 74 (the asterisk indicates invariant residues).

On the N-terminal side of the first helix the zinc-binding site is formed between a 13-residue loop (between the two pairs of cysteines) and the end of the helix. Contacts between the hydrophobic residues, Tyr 17 and Trp 22, define the base of this loop. The N terminus folds back to form a second loop in which there are hydrophobic interactions between Thr 4, Tyr 6 and Tyr 13. The structure is consolidated through interaction of the tips of both of these loops with other regions of the protein. In one loop, Cys 10 interacts with Ile 51, and in the other, both Val 21 and Tyr 19 interact with the interhelix hydrophobics.

On the N-terminal side of the second helix there is a short five-residue loop between Cys 43 and Cys 49 and a nine-residue loop between Cys 49 and Cys 59. The base of both of these loops is defined by coordination to the zinc ion. There are also several contacts between Ile 51 and both the second helix and Cys 10.

What can we understand about the nature of the specific interaction of the ERDBD with DNA from knowledge of the structure of the monomer? Comparison of the palindromic response elements of the oestrogen, glucocorticoid and thyroid hormone receptors (ERE, GRE and TRE, respectively; Fig. 4a) indicates that the protein must discriminate both between different half-site sequences (ERE versus GRE) and different half-site spacings (ERE versus TRE).

Several mutagenesis studies have led to the identification of three residues that are necessary and sufficient to cause differential recognition of the ERE and GRE (Glu 25, Gly 26 and Ala 29 versus Gly, Ser and Val)^{22,23,24}. These residues are exposed to solution at the N terminus of the first helix (Fig. 3c). On the same face of the helix there are also two conserved lysines and an arginine residue, suggesting that this helix is a DNA recognition helix.

Further mutagenesis studies²³ indicate that the five residues, Pro 44 to Gln 48, in the second zinc-binding site, are involved in discrimination of half-site spacing (between the ERE and TRE). This suggests that these residues are involved in the spacing of the two monomers when complexed with DNA. In the structure of the monomer these residues are located on the surface of the molecule at about 19 Å from the recognition helix. This is consistent with the recognition helices of each monomer binding in adjacent major grooves of the DNA (separated by 34 Å). Figure 4b shows a schematic model to indicate how two DNA-binding domains can form a dimer on binding with the palindromic response element. In this model, the conserved basic side chains, Arg 56 and Lys 57, are in a suitable position

to interact with the phosphate backbone of the DNA in the minor groove between the two DNA-binding domains. Indeed, when these residues are mutated to glycines, DNA binding is virtually abolished¹⁹. It seems that the protein in this region is relatively flexible in solution. If this were to become ordered in the dimer-DNA complex, it could account for the highly cooperative binding observed in bandshift assays (J.W.R.S. *et al.*, manuscript in preparation).

During preparation of this manuscript a paper was published describing the NMR structure of the glucocorticoid receptor DNA-binding domain²⁵. This structure seems to be similar to that of the oestrogen receptor, but there are some differences in the orientation of the nine-residue loop and hence in the potential for binding to the minor groove of DNA.

The conservation of structurally important amino acids in related proteins suggests that this 'double zinc-helix' structural motif is common to other members of the superfamily of steroid hormone receptors and is thus a general structure for protein-DNA recognition. □

Received 16 August; accepted 19 October 1990.

1. Yamamoto, K. R. *A. Rev. Genet.* **19**, 209-252 (1985).
2. Green, S. & Chambon, P. *Trends Genet.* **4**, 309-314 (1988).
3. Evans, R. M. *Science* **240**, 889-895 (1988).
4. Green, S. & Chambon, P. *Nature* **325**, 75-78 (1987).
5. Kumar, V. *et al. Cell* **51**, 941-951 (1987).
6. Kumar, V. & Chambon, P. *Cell* **55**, 145-156 (1988).
7. Klug, A. & Rhodes, D. *Trends biochem. Sci.* **12**, 464-469 (1987).
8. Evans, R. M. & Hollenberg, S. M. *Cell* **52**, 1-3 (1988).
9. Freedman, L. P. *et al. Nature* **334**, 543-546 (1988).
10. Berg, J. M. *Cell* **57**, 1065-1068 (1989).
11. Lee, M. S. *et al. Science* **245**, 635-637 (1989).
12. Kleit, R. E., Herriott, J. R. & Horvath, S. J. *Proteins* **7**, 215-226 (1990).
13. Neuhaus, D., Nakeseko, Y., Nagai, K. & Klug, A. *FEBS Lett.* **262**, 179-184 (1990).
14. Picard, D. & Yamamoto, K. R. *EMBO J.* **6**, 3333-3340 (1987).
15. Studier, F. W. & Moffatt, B. A. *J. molec. Biol.* **189**, 113-130 (1986).
16. Wüthrich, K. *NMR of Proteins and Nucleic Acids* (Wiley, New York, 1986).
17. Clore, G. M. & Gronenborn, A. M. *Crit. Rev. biochem. molec. Biol.* **24**, 479-564 (1989).
18. Severne, Y., Wieland, S., Schaffner, W. & Rusconi, S. *EMBO J.* **7**, 2503-2508 (1988).
19. Hollenberg, S. M. & Evans, R. M. *Cell* **55**, 899-906 (1988).
20. Nilges, M. in *Brünger, A. T. in X-PLOR Version 2.1. User Manual* (Yale University Press, 1990).
21. Brünger, A. T. *X-PLOR Manual* (Yale University Press, 1988).
22. Mader, S., Kumar, V., de Verneuil, H. & Chambon, P. *Nature* **338**, 271-274 (1989).
23. Umesono, K. & Evans, R. M. *Cell* **57**, 1139-1146 (1989).
24. Danielsen, M., Hinck, L. & Ringold, G. M. *Cell* **57**, 1131-1138 (1989).
25. Härd, T. *et al. Science* **249**, 157-160 (1990).
26. Shaka, A. J., Lee, C. J. & Pines, A. J. *magn. Reson.* **77**, 274-293 (1988).

ACKNOWLEDGEMENTS. We are indebted to M. Nilges, D. Norman and I. Campbell for giving us the YASAP protocol and advice in the use of XPLOR, Y. Nakeseko for advice in the preparation of protein samples, K. Butler and J. Butler for assistance with atomic absorption spectroscopy and analytical ultracentrifugation experiments, respectively, G. Jacobs for advice in molecular graphics programs, W. Sundquist, L. Fairall and P. Powell for critical comments on the manuscript and A. Klug for continued interest and support. J.W.R.S. holds MRC Studentship.

Micropreparative LC of proteins

E. C. Nice

The use of short microbore HPLC columns with different selectivity for the purification of proteins and peptides at the picomole level is discussed.

THE synergism between protein chemistry and molecular biology enables protein amino acid sequence data to be translated into the corresponding oligonucleotide, which can then be used to isolate the gene encoding the protein of interest. Using recombinant DNA techniques, the gene can then be used to produce amplified quantities of the protein itself, enabling detailed structure-function studies or clinical trials to be performed on proteins that, hitherto, could be isolated only from biological samples in microgram quantities.

It is now possible to obtain amino acid sequence information at the low picomole level. Such studies rely upon the ability to purify nanogram-to-microgram quantities of proteins or peptides with high yield, and in a form suitable for amino acid sequence analysis. Samples should be in small volumes (20–100 μ l) to facilitate loading onto the sample disc of the sequenator, and free of salts and other contaminants that could interfere with the Edman chemistry. At such low levels, techniques such as lyophilization, dialysis and organic solvent precipitation should be avoided because they are invariably associated with unacceptably high losses^{1,2}. The ability to manipulate proteins and peptides within the confines of a chromatographic column and, ultimately, recover them in small volumes and in high yield would therefore appear to be advantageous.

For columns of varying dimensions, operated under identical linear flow velocities, it can be shown that the eluent peak volumes observed are inversely proportional to the square of the column diameter and directly proportional to the column length³. Short (< 10 cm), microbore (1–2-mm i.d.) columns, operated at flow rates of 50–100 μ l min⁻¹, are ideal for the purification, concentration and buffer exchange of samples^{4–7}. Such columns show excellent resolution of complex protein mixtures^{4,6,7}. The small peak volumes obtained (20–100 μ l), with the concomitant high protein concentrations, means that nanogram quantities of protein are readily detectable^{4,6,7}.

In practice, this high sensitivity means that sample recovery can be monitored and chromatographic conditions optimized using a small portion of the sample before committing the bulk of the material. Even at these very low levels, the purified components can be recovered

and added back to the bulk of the preparation in order to avoid unnecessary losses. It has been shown^{4,7} that such columns are capable of chromatographing 10–20 μ g of protein under optimal conditions. Such quantities are ideally suited for the preparation of samples for microsequence analysis and for the final stages of the multidimensional purification of proteins present at trace levels in bulk biological samples.

If the full benefit of such microcolumns is to be realized, it is essential that large sample volumes can be loaded. As proteins have very high capacity factors on interactive supports (for example, reversed phase (RP), hydrophobic interaction, ion exchange, or affinity matrices) they can be concentrated onto such supports by trace enrichment using suitable sample loading conditions. The resultant protein concentrations are sufficiently high to allow small aliquots of the recovered samples to be analysed by capillary zone electrophoresis, SDS-PAGE or plasma desorption mass spectroscopy to ascertain sample homogeneity.

While the use of micropreparative HPLC systems to micromanipulate proteins and peptides before sequence analysis is gaining in acceptance, most examples to date have used RP-supports. This has, no doubt, been due to the availability of such columns and their ability to separate, with excellent resolution, many proteins and peptides. The use of volatile solvent systems (such as, TFA/CH₃CN gradients) facilitates application of the purified material to the sequenator. In addition, the small peak volumes are compatible with the direct sequence of salt-containing samples, allowing samples to be sequenced directly when they have been re-chromatographed on RP-columns using an alternative mobile phase (such as, 0.9% NaCl, pH 2.1/CH₃CN) to modulate selectivity and, hence, obtain homogeneous preparations.

'Inverse gradient' technique

The range of proteins that can be chromatographed on RP-supports has been further extended using a novel 'inverse gradient' technique^{8,9}. This technique is based on observations that certain normal¹⁰ and RP-supports^{11,12} retain proteins at high (> 60%) concentrations of organic solvent. Typically, the RP-supports that display this phenomenon have small pore

sizes (60–120 Å) with large surface areas (200–400 m² g⁻¹) and high carbon loadings (10–18%). Proteins can be retained on such supports at high concentrations of organic solvent (90–100%) and then recovered in high yield (routinely > 85%) by the simultaneous addition of an ion-pairing reagent, while running a gradient to a lower organic solvent concentration. This technique has proved extremely useful for the recovery and desalting of proteins from SDS-PAGE electroeluates^{8,9} and from Coomassie blue-stained proteins electroblotted from gels onto PVDF membranes¹³ in a form that is suitable for sequence analysis. Interestingly, several proteins (such as, insulin receptor and gastrin binding protein), which cannot be recovered from RP-supports using gradients of increasing concentrations of organic solvent, can be recovered in high yield when using inverse gradients^{8,9}.

Alternative selectivity

Nevertheless, many proteins are not compatible with reversed-phase methodologies because they are either irreversibly retained by the support or denatured by the solvent system in use, thus precluding the use of specific bioassays to monitor purification. The use of microcolumns with alternative selectivity would therefore greatly extend the applicability of the technique. A range of columns is now available that can be used with the SMART System (Pharmacia LKB, Uppsala, Sweden), a micropurification system with an associated microfraction collector. It incorporates an 'expert' system (SMART Assistant) to assist in the development of purification protocols.

Protein recoveries have been studied using a technique in which a protein is injected onto the column, recovered, diluted with primary buffer (containing low concentrations of detergent, such as 0.02% Tween 20, to minimise losses by non-specific adsorption), and re-injected back onto the same column. The repetitive recoveries observed with both anion (Mono Q PC 1.6/5) and cation (Mono S PC 1.6/5) exchange columns were about 90%¹⁴. The separation of a set of protein standards on a micropreparative anion exchange column is shown in Figure 1a. Here, the column was eluted with a NaCl gradient in which the pH had been controlled by the dropwise addition of 0.1 M phosphate buffer (final concentration

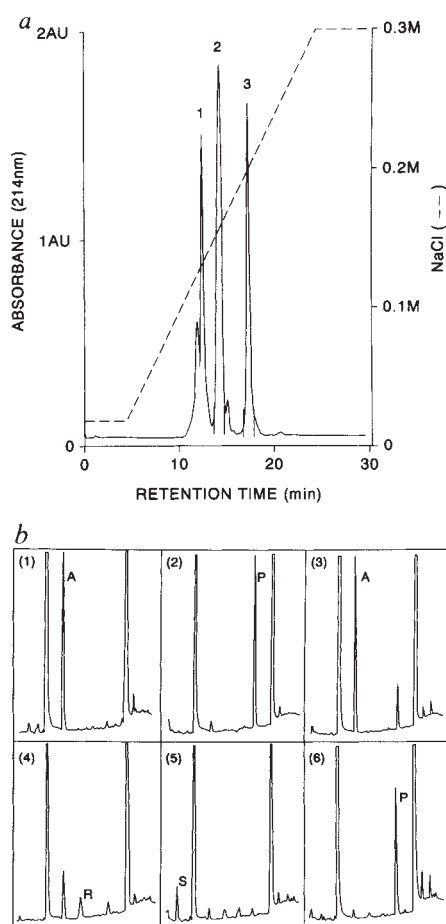


FIG. 1 Purification of proteins on micropreparative anion exchange columns for N-terminal sequence analysis. *a*, Transferrin (peak 1: 10 μ g), recombinant human epidermal growth factor (peak 2: 2 μ g) and recombinant human granulocyte macrophage colony stimulating factor (hGM-CSF) (peak 3: 10 μ g) were separated on a Pharmacia LKB Mono Q PC 1.6/5 column using a linear 20-min gradient between 20 mM NaCl (pH 6.8) and 300 mM NaCl (pH 6.6). The flow rate was 100 μ l min⁻¹. The purified proteins were recovered for N-terminal sequence analysis. *b*, The first six cycles of the generated sequence data for the hGM-CSF (peak 3) are shown. The anticipated sequence was clearly identifiable, with an initial sequence yield of 40% (U. Hellman, personal communication).

0.067 mM). The small peak volumes and virtual lack of any buffer components that could interfere with the Edman chemistry or the subsequent analysis of the generated Pth-amino acids, permitted the separated components to be loaded directly onto the sample disc of the sequencer, and the amino acid sequence obtained (see Fig. 1*b*). In addition, the tandem use of ion exchange and reversed-phase purification has been shown to simplify the purification of complex mixtures of peptides from enzymic digests^{7,14}.

Desalting microcolumns

Even if one is successful in purifying a protein to homogeneity, the N-terminal sequence data obtained may not be suit-

able (because of a blocked N-terminus, degenerate or uninterpretable sequence) for the generation of an oligonucleotide probe to isolate the gene of interest. In such cases, internal sequence is usually obtained from peptides generated by the enzymic or chemical fragmentation of the protein following reduction and alkylation. The generation and purification of peptides by this route is not trivial, and presents the researcher with a further set of manipulative and separative problems. In particular, the desalting of the reduced and alkylated proteins before enzymic digestion is often problematical¹⁵. The use of size exclusion or desalting columns would appear to be the method of choice. However, as such columns are not interactive (that is, trace enrichment is not possible), sample volumes become limiting and, if small peak volumes compatible with subsequent micromanipulations are required, column dimensions must be carefully 'tuned' to suit the sample volume applied. The use of such a column (Fast

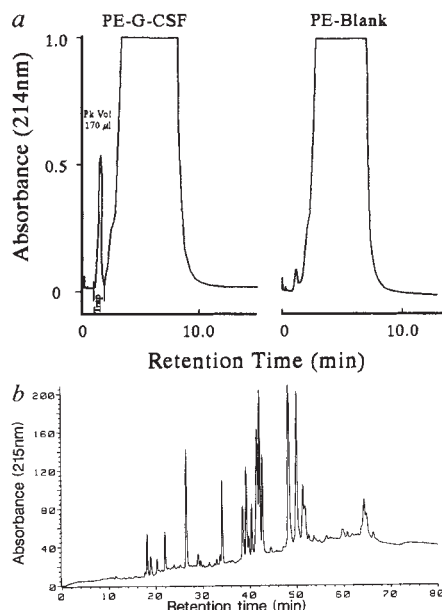


FIG. 2 Use of micropreparative desalting columns to recover pyridine ethylated granulocyte colony stimulating factor (G-CSF) for *Staphylococcus aureus* V8 digestion. Recombinant hG-CSF (2.5 μ g) was dissolved in 90 μ l of 0.25 Tris buffer (pH 8.5) containing 6 M GuHCl and 1 mM EDTA. DTT (50 μ g) was added and the reaction mixture flushed with argon and incubated for 3 hr at 25 °C in the dark, before adding vinyl pyridine (2 μ l) and incubating for a further 1 hr. *a*, The reduced and pyridine ethylated G-CSF was separated from other reagents on a Fast Desalting PC 3.2/10 column, using 0.1% NH₄HCO₃/0.02% Tween 20 as the mobile phase at a flow rate of 190 μ l min⁻¹. A reagent blank is shown for comparison. *b*, The recovered protein (peak vol. 170 μ l) was digested for 18 hr at 37 °C with 1 μ g *Staph. aureus* V8 protease and the resultant peptides separated on a micropreparative RP-HPLC column (μ RPC 3.2/3) using a linear 60-min gradient between 0.15% TFA and 60% CH₃CN/0.125% TFA. The flow rate was 100 μ l min⁻¹.

Desalting PC 3.2/10 packed with Sephadex G-25 Superfine) for the desalting and buffer exchange of a haemopoietic growth factor (granulocyte colony stimulating factor: G-CSF) following reduction and pyridine ethylation before enzymic digestion is shown in Figure 2.

In conclusion, the use of micropreparative columns is advantageous for the purification of microgram-to-nanogram quantities of proteins or peptides. The small peak volumes and resultant increase in sample concentration help reduce losses by non-specific adsorption, lead to increased sensitivity of detection and facilitate subsequent manipulations. The availability of a range of microcolumns with alternative selectivity makes the technique applicable to most purification problems. □

Edouard C. Nice is at the Ludwig Institute for Cancer Research, P.O. Royal Melbourne Hospital, Victoria 3050, Australia. For more information, fill in reader service number 100.

1. Esch, F.S. *Analyt. Biochem.* **136**, 39–47 (1984).
2. Hunkapiller, M.W. & Hood, L.E. *Science* **219**, 650–659 (1983).
3. Kucera, P. *J. Chromatogr.* **198**, 93 (1980).
4. Nice, E.C., Lloyd, C.J. & Burgess, A.W. *J. Chromatogr.* **296**, 153–170 (1984).
5. Moore, R.M. & Walters, R.R. *J. Chromatogr.* **317**, 119–128 (1984).
6. Nice, E.C., Grego, B. & Simpson, R. *Biochem. Int.* **11**, 187–195 (1985).
7. Simpson, R.J., Moritz, R.L., Begg, G.S., Rubira, M.R. & Nice, E.C. *Analyt. Biochem.* **177**, 221–236 (1989).
8. Simpson, R.J., Moritz, R.L., Nice, E.C. & Grego, B. *Eur. J. Biochem.* **165**, 21–29 (1987).
9. Simpson, R.J. & Moritz, R.L. *J. Chromatogr.* **474**, 418–423 (1989).
10. Rubenstein, M. *Analyt. Biochem.* **98**, 1–7 (1979).
11. Bij, K.E., Horváth, Cs., Melander, W.R. & Naham, A. *J. Chromatogr.* **203**, 65–84 (1985).
12. Hearn, M.T.W. & Grego, B. *J. Chromatogr.* **255**, 125–136 (1981).
13. Ward, L.D., Reid, G.E., Moritz, R.L. & Simpson, R.J. in *Current Research in Protein Chemistry* (ed. Villafrañe, J.J.) (Academic, New York, 1990).
14. Nice, E.C. & Simpson, R.J. *J. Pharmaceut. Biomed. Anal.* **7**, 1039–1053 (1989).
15. Simpson, R.J. & Nice, E.C. in *The Use of HPLC in Receptor Biochemistry* (ed. Kervilave, A.R.) 201–244 (A.R. Liss, New York, 1989).

ADVERTISEMENT

**Teflon[®]
Coated
and**

**Bare Wire from
MEDWIRE**

Medwire provides the highest quality platinum, platinum alloy, silver and stainless steel wire in both single and multistrand configurations.

For prompt delivery or additional information, please contact:

MEDWIRE, an affiliate of Sigmund Cohn Corp.
121 So. Columbus Ave. Mount Vernon, NY 10553
Tel: (914) 664-5300 Fax: (914) 664-5377

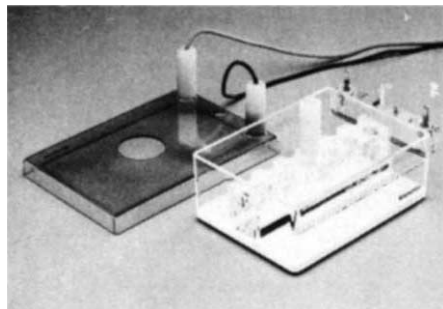
*Reg. TM of E.I. DuPont & Co.

Reader Service No.25

Separation technology sources

DNA extraction cartridges, a dual-purpose electrophoresis/electroblotter unit, and a cytocentrifuge that makes light work of preparing cell monolayers on microscope slides — innovative ideas for separation devices.

USING the HSB-Elutor it is possible to perform six **electrophoretic elutions of nucleic acids** in parallel, without the loss of material by adsorption to membranes, says Biometra (*Reader Service No. 101*).



Run six elutions in parallel with the HSB-Elutor from Biometra.

Macromolecules initially migrate within a solution of low ionic strength but, on meeting a solution of high ionic strength located in the V-shaped channel between the buffer reservoirs, migration virtually stops and the macromolecules concentrate. Typically, the £475 (UK) HSB-Elutor produces end volumes of 100 µl.

Pharmacia has introduced the Monitor UV-M II **fixed-wavelength UV detector** for FPLC, high performance and standard chromatography (*Reader Service No. 102*). The detector module can be positioned vertically or horizontally, which enables the user to tailor the system to suit specific requirements. It features touch-sensitive controls for setting sensitivity, baseline adjustment, event mark and autozero. The small optical unit can be placed directly after the column, minimizing the dead volume between the column and the monitor, says Pharmacia. Eight filters and interchangeable mercury and zinc lamps provide detection over a range of 214 to 546 nm.

Extracting with ease

The new SpinBind **microcentrifuge cartridges** from FMC BioProducts are designed to purify DNA in about 15 minutes (*Reader Service No. 103*). SpinBind DNA extraction units contain a microporous plastic/silica sheet that binds DNA in the presence of high concentrations of chaotropic salts, says FMC BioProducts. After washing, DNA can be recovered in a low ionic strength buffer and used directly for enzymatic manipulation and cloning purposes. SpinBind DNA extraction units cost \$110 (US) for 50 units.

ImmunoSelect **recombinant Protein G membrane discs** from Life Technologies are designed to speed up IgG purification from large volumes and dilute samples, while eliminating the need for packed columns (*Reader Service No. 104*). Recombinant Protein G binds only IgG-isotype antibodies and, says the company, recognises more species than Protein A. It binds well to mouse IgG-isotype and rat IgG antibodies. Recombinant Protein G does not bind to serum albumin, IgM, IgA, IgE or IgD. The user equilibrates the disc, loads the sample, washes, and then elutes the purified IgG. To speed up the process when using large volumes, a peristaltic pump can be used.

In a spin

A choice of three acceleration rates and the facility to store up to nine programmed routines in the non-volatile memory are features of the Cytospin 3 **cytocentrifuge** from Shandon Scientific (*Reader Service No. 105*). The Cytospin 3



Produce cell monolayers in moments with the Cytospin 3.

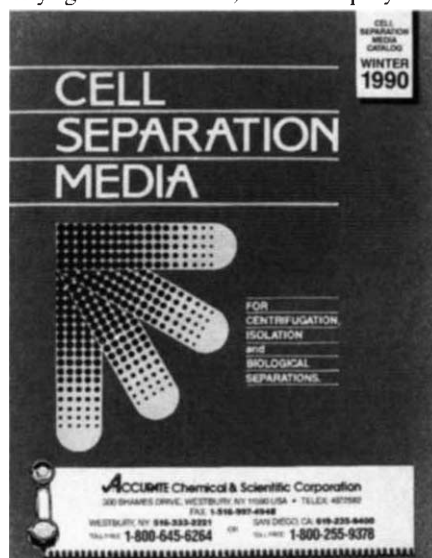
cell preparation system can be used to prepare a cell monolayer on a small, well-defined zone of a microscope slide, while the suspension medium is absorbed onto a filter card. The cytocentrifuge has applications in cytology, haematology and microbiology, where cell preparation procedures that preserve the integrity of fragile cells are required. Cytospin 3 can operate at speeds of between 200 and 2,000 r.p.m. and can be programmed for run times of between one and 99 minutes. Cell preparation procedures can be standardized and operator time saved by using the instrument's nine-programme memory. As a safety measure, Shandon supplies an autoclavable sealed head to prevent aerosol contamination.

The Marathon Model 13K/M **microcentrifuge** from Fisher Scientific features a

rugged heavy-gauge steel construction and an autoclavable 24-place angle rotor for 1.5-ml tubes, although 0.4- and 0.5-ml tubes can be accommodated using optional reduction plates (*Reader Service No. 106*). Samples can be processed at speeds of up to 12,000 r.p.m. (13,200 g) and for run times of 0–30 minutes in one-minute increments. Other design features of the \$1,795 (US) microcentrifuge include an air-draft design for keeping samples cool at high speeds, a momentary spin button for short spin cycles, and an LED readout of elapsed time. For safety purposes, the 13K/M can be fitted with a gasketed snap-on rotor lid to contain aerosol contamination. The 13K series also includes clinical and haematocrit versions.

Media mix

Details of new lines in **cell separation media** for centrifugation, isolation and biological separation can be found in the new 20-page catalogue from Accurate Chemical and Scientific (*Reader Service No. 107*). New product lines include density gradient media, defined polymers



Media by mailorder.

and relevant instrumentation. A handy cross-reference guide is included in the catalogue to help the researcher match products to particular applications.

Two new ready-to-use SequaGel **sequencing gels** are available from National Diagnostics, which are designed for improved reproducibility and convenience of casting while reducing exposure to toxic acrylamide dust (*Reader Service No. 108*). SequaGel-6 is suitable for the

preparation of six per cent sequencing gels that can be used to sequence DNA fragments from 60–150 bases. SequaGel-8 makes eight per cent gels and is designed for the sequencing of DNA fragments in the 40–100-base range. SequaGel solutions already contain acrylamide, bis-acrylamide, urea, TBE buffer and TEMED; the user just adds ammonium persulphate before pouring the gel. A gaseous inhibitor is added to stabilise the solutions. In this form, the monomer is prevented from the self-polymerization to which solid acrylamide is susceptible, says the company. SequaGel solutions are available in 500-ml and 1-litre quantities.

Amresco has introduced a range of **pre-mixed liquid acrylamide** products that eliminate the need to handle hazardous powdered acrylamide (*Reader Service No. 109*). Several formulations are available: in addition to the standard 40% (w/v) acrylamide solution and a standard 2% bis-acrylamide solution, Amresco offers three different 40% solutions containing acrylamide/bis-acrylamide in



Ready-to-use liquid acrylamide.

ratios of 19:1, 29:1 and 37.5:1. By simple dilution with either deionized water or buffer, the user can prepare gels of any concentration between zero and 40%. Unlike other products, the liquid acrylamide does not contain gaseous stabilisers. When refrigerated, liquid acrylamide remains stable for up to a year, says Amresco. All solutions are available in 500-ml or bulk quantities.

Zymogram **pre-cast mini-gels** from Novex have key applications in the study of tumour metalloproteins and their role in malignancy (*Reader Service No. 110*). The zymogram gel, which consists of a 10% PAGE gel incorporating gelatin as a substrate, can be used to detect and characterize metalloproteins, collagenases and other proteases. After running and processing, proteases are visible as clear bands against a blue background, says the company. Novex also supplies buffer solutions for running the gel and for gel development.

Electrophoresis: in focus

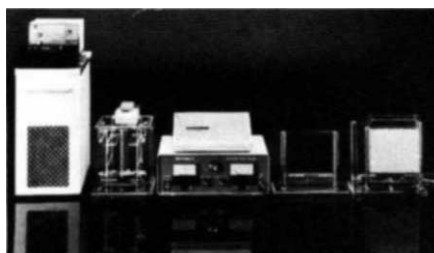
The compact Model 3850 **capillary electrophoresis system** from Isco supports a



Isco's capillary electrophoresis system with femtomole sensitivity.

full range of CE modes, including micellar electrokinetic capillary chromatography, as well as conventional free-solution methods with coated and non-coated capillaries (*Reader Service No. 111*). The instrument features a 30-kV reversible-polarity power supply, and 190–360-nm variable detection with femtomole sensitivity. The split-flow injector uses an HPLC-style syringe that is designed to provide reproducible nanolitre injections from microlitre samples. The system can be used for the separation and analysis of proteins, peptides, amino acids, DNA fragments and pharmaceutical compounds.

Beckman has developed a pulsed-field electrophoresis system for DNA mapping called GeneLine II that uses **transverse alternating field electrophoresis** (TAFE) technology to separate large fragments in straight lanes and with high resolution (*Reader Service No. 112*). A special feature of the system is a "touch-free" gel handling where gels are cast directly in the acrylic frame. This set-up is designed to support and protect even "soft" gels with agarose concentrations as low as 0.6%. Temperature control is maintained by the



GeneLine II uses TAFE technology.

buffer cooler and the vertical positioning of the gel. The complete GeneLine II system consists of the electrophoresis chamber, stain and circulating rinse tanks, power supply, controller, buffer cooler and pump, 20-tooth analytical comb and preparative comb, "touch-free" gel holder, gel mould and reagents.

Hybaid is offering a complete line of **horizontal gel electrophoresis tanks** for the separation of DNA and RNA in agarose gels, which feature three-point levelling and a no-tape, Easi-Cast gel casting system (*Reader Service No. 113*). Three sizes are available: £159 (UK) mini (7 × 8

cm), £190 (UK) midi (12 × 14 cm) and £285 (UK) maxi (20 × 25 cm). All are fitted with platinum electrodes that are protected against corrosion and lined with Teflon to provide an even electrical field. To cast a gel in the mini and midi systems, the gel tray is placed sideways in the chamber before pouring the agarose. When the gel is set, the tray is then turned back to the running position. Stated applications for the maxi system include large-scale screening, RFLP analysis and DNA fingerprinting.

Perform electrophoresis and electrotransfers using the same piece of apparatus with the DuoLink **combined electrophoresis and blotting module** from AT Biochem (*Reader Service No. 114*). Users can run two gels simultaneously and, by changing the insert, perform electrotransfers in one hour or less, says the

ADVERTISEMENTS

Synthetic Oligonucleotides

Custom made Probes and Primers,
Synthesised, purified and shipped in 3 days

£4.00 per base

(All inclusive)

0.2 µM scale (approx. 500 µg)

Modified bases, aminoalkylation,
data base analysis and full design facilities

Molecular Medicine Unit,
King's College School of Medicine & Dentistry,
Denmark Hill, London SE5 8RX

Tel. XX-44-(0)71 - 326 3126

FAX XX-44-(0)71 - 733 3877

Reader Service No.56

CUSTOM-MADE & PURIFIED DNA PEPTIDES

\$8/base

50–100 µg

40 hour delivery

Quotes available

50–300 mg

1–3 wks delivery

GUARANTEED QUALITY & QUANTITY

V.G. CORP.

1200 Wonderland Rd. S
Bldg #7, Unit #2
London, Ontario
Canada N6L 1A8

CALL OR FAX: (519) 652-3758

Reader Service No.29



Custom DNA Synthesis
Probes • Primers • Oligo's

For a fast efficient
Cost effective service

BUY SEVERN BIOTECH

Tel: 0562 825286 Fax: 0562 825284

Unit 23, Hoobrook Enterprise Centre,
Worcester Road, Kidderminster DY10 1HY

Reader Service No.7

ADVERTISEMENTS

TIB MOLBIOL BERLIN
CUSTOM DNA SYNTHESIS

5,- DM per base

DM 95,- per oligo

FAX (0049) 30 6261960

Reader Service No.80

TIB MOLBIOL BERLIN
CUSTOM DNA SYNTHESIS

ANTGCCAT
ANTISENSE

5'-PSORALEN and more.....

TIB MOLBIOL BERLIN
FAX (0049) 30 6261960

Reader Service No.22

ATTENTION MATHEMATICIANS

Available now from booksellers:
**FERMAT'S LAST THEOREM
PROVED & AWARD OFFERED
FOR REFUTATION** — by Shafi
Ahmed. Price £8.95 ISBN 0 948331
054. Or directly from: **Oaktown
Publications, 100 Western Ave.
London W3 7TX.** Orders from
abroad welcome. U.S. \$20 includ-
ing Postage.

Reader Service No.23

NIMA
TECHNOLOGY

Langmuir Troughs
Thin Film Balances
Tensiometers
Bilayer Membrane Systems

For all your requirements at the gas-liquid interface.

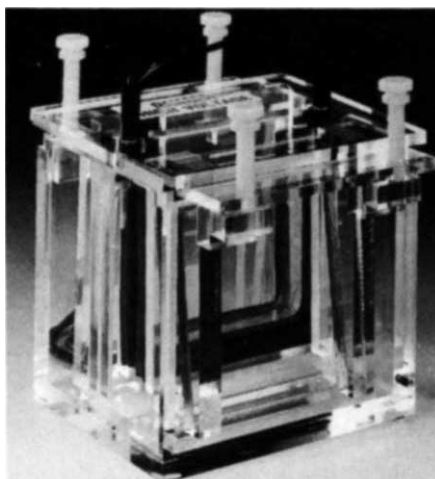
CTC Technologies Inc.
7925A North Oracle Rd.
Tucson, AZ 85704, USA
Tel: 800 CTC TECH.

Nima Technology Ltd.
The Science Park,
Coventry, England.
Tel: 0203 419457.

Reader Service No.24

**Copies of articles from this
publication are now available from
the UMI Article Clearinghouse.**

Mail to: University Microfilms International
300 North Zeeb Road, Box 91 Ann Arbor, MI 48106



Dual-purpose electrophoresis/blotter unit.

company. The \$600 (US) DuoLink module is compatible with all vertical mini-format techniques and can accommodate 10 × 10-cm glass gel plates or cassettes. The apparatus is supplied with two sets of plates, 1.5-mm combs and spacers.

Tools for TLC

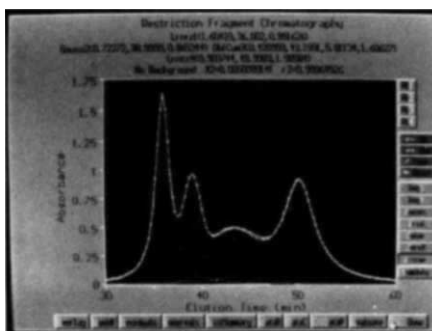
Whatman has developed a new line of **preadsorbent TLC plates** that are designed to provide enhanced sensitivity over conventional silica plates, allowing even dilute samples to be analysed accurately and reproducibly (*Reader Service No. 115*). The TLC plates feature a 2–3-cm band of non-adsorbing, inert material to which the samples are applied. The mobile phase then compresses the samples into narrow bands at the interface with the analytical layer. Thin layer chromatography separation then proceeds as normal. As the preadsorbent band is 20 per cent thicker than the analytical area, Whatman says that sample volumes of up to 20 µl can be applied. The preadsorbent TLC plates are available with a range of analytical phases: standard silica gel, high performance silica gel, or C₁₈ reversed-phase. In addition, the plates can be supplied in either a plain or pre-channelled format. The latter is useful for preventing cross-contamination.

Camlab is distributing the Desaga AS30 **autospotter for TLC plates**, which has both spotting and streaking capabilities (*Reader Service No. 116*). The operating principle of the TLC applicator is that a stream of gas carries the sample from the cannula tip onto the TLC plate; this prevents damage to the layer and provides a means of non-contact application in the form of a line or a dot. The instrument accepts TLC plates and sheets of up to 200 × 200 mm and path lengths of between 0 and 100 mm. All the parameters — plate size, number, length and distance of the paths, the volumes applied, and the rate of application — for the application of up to 30 samples are entered via the

keyboard. Up to ten different methods can be stored in the instrument's memory. The TLC applicator, which costs about £3,750 (UK), can be interfaced to an auto-sampler, via a serial interface to form a fully automatic application system.

Software options

PeakFit, Jandel Scientific's latest **scientific and engineering software package**, is designed for separating and analysing multiple, overlapping peaks or functional forms (*Reader Service No. 117*). The software has been developed as an aid to scientists and engineers working with spectroscopic peaks, chromatography profiles, chemical and pharmacological kinetics data, waveform and engineering data, statistical distributions and transition functions. PeakFit offers both non-



PeakFit solves curve-fitting problems.

linear curve-fitting and peak quantification (amplitude, area, centres). The user selects a function for the data and PeakFit automatically lays down a best guess for the fit. The amplitude, centre location and width of the peak on the screen can then be manipulated. Once positioned, a single keystroke initiates the iteration process, which is shown graphically until convergence. The \$595 (US) PeakFit program features pull-down menus, mouse support and the ability to operate in the Microsoft Windows environment.

MACScience Solutions announces the availability of the Macintosh-based **Peaks chromatography analysis system** for the acquisition, analysis and display of chromatography data (*Reader Service No. 118*). The system consists of a \$2,995 (US) four-channel analog-to-digital converter and the \$995 (US) Peaks chromatography analysis software program. All of the commonly used features — baseline correction, injection times and peak detection/integration — are managed through the use of screen controls. Runs can be compared using the overlay function. □

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain additional information about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.